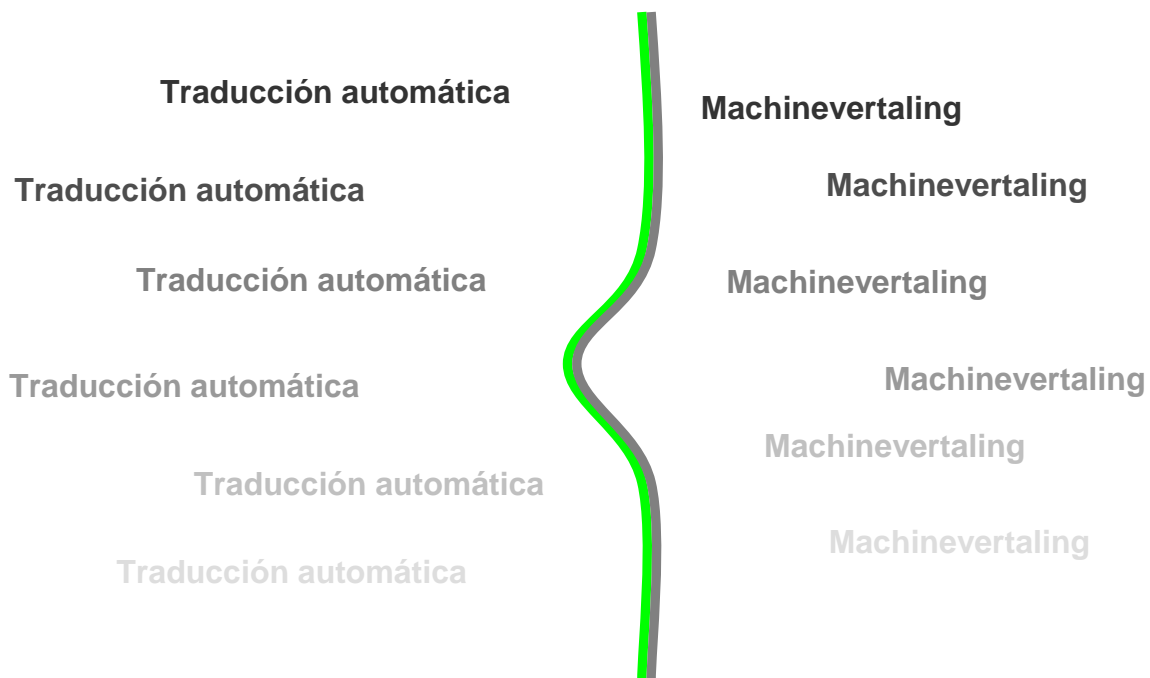


EL ALCANCE DE LA TRADUCCIÓN AUTOMÁTICA

UN ESTUDIO DE LA TRADUCCIÓN AUTOMÁTICA DEL PAR DE LENGUAS ESPAÑOL-NEERLANDÉS



“El lenguaje ha de ser matemático, geométrico, escultórico.

La idea ha de encajar exactamente en la frase, tan exactamente que no pueda quitarse nada de la frase sin quitar eso mismo de la idea.”

José Martí (1853-1895)

Noviembre de 2010

Universiteit Utrecht

Máster de Traducción

Lengua y cultura española

Estudiante: Rosanne Tertoolen
3180530

Supervisora: Dr. D. Nieuwenhuijsen

Código 200401048

Índice

Introducción

1	Introducción a la traducción automática	6
	1.1 El valor de la traducción automática	6
	1.2 ¿Qué es la traducción automática?	8
	1.3 La evolución de la traducción automática e investigación	12
	1.4 Conclusión	16
2	La complejidad de la traducción automática	18
	2.1 Limitaciones y problemas	19
	2.2 Ambigüedad	23
	2.3 Conclusión	27
3	Metodología	28
	3.1 Objetivo del estudio	28
	3.2 La evaluación de la traducción automática	29
	3.3 El método seleccionado	30
	3.4 La selección de los textos	33
	3.5 Conclusión	36
4	¿Cuál es el alcance?	37
	4.1 La evaluación lingüística	37
	4.2 Ambigüedades en los textos	48
	4.3 Las valoraciones de los expertos	54
	4.4 La calidad después de 'post-editar'	61
	4.5 Conclusión	65
	Conclusión	6

Referencias

Anexos

Introducción

El ser humano es capaz de lograr cosas que hasta hace poco tiempo se antojaban imposibles. Hemos conseguido aterrizar en la Luna, manipular el ADN y clonar animales. Sin embargo, y pese a estos innegables avances, aún no hemos logrado automatizar una lengua. En los comienzos de la investigación en la traducción automática, los investigadores creían en la idea de lograr un *fully automatic high quality translation* (FAHQT) en un par de años. Después de 60 años de investigación casi nadie cree que un día logremos esta traducción automática de alta calidad. Incluso hay muchos que opinan que apenas se ha hecho progreso en los 60 años de investigación. Gracias a Internet hay una gran cantidad de información fácilmente accesible para todo el mundo. Sólo hay una barrera, que a primera vista parece minúscula: la barrera del idioma. Si uno no pertenece a una determinada comunidad lingüística o si no ha aprendido su idioma, no tiene acceso a la información. A este respecto, el valor de la traducción automática es de gran importancia.

Siendo estudiante del *Máster de Traducción*, he notado que la traducción automática empieza a incorporarse en el trabajo de los traductores. No sólo estudiantes del *Máster de Traducción* trabajan con la ayuda de la traducción automática, sino también traductores profesionales. Con la llegada del ordenador y luego el Internet, la profesión de la traducción ya ha cambiado mucho; la máquina de escribir fue reemplazada por el ordenador con el programa de tratamiento de textos y con el controlador de ortografía, los diccionarios en papel fueron reemplazados por los diccionarios electrónicos y la biblioteca fue reemplazada por la biblioteca virtual: el Internet. Es obvio que la profesión del traductor ya ha experimentado cambios enormes, y tal vez estemos en los principios de un nuevo cambio radical: la incorporación de la TA en el trabajo de los traductores. Por motivo de estos desarrollos nos pareció interesante y relevante investigar la utilidad de la TA a la hora de

traducir un texto. El objetivo del estudio será: *Determinar el alcance actual de la traducción automática en cuanto al par de lenguas español-neerlandés.*

No pretendemos dar soluciones para mejorar los sistemas, sino que, a través de evaluaciones de textos traducidos automáticamente, intentamos identificar los puntos fuertes y débiles.

En el primer capítulo expondremos los diferentes sistemas de la traducción automática y discutimos los desarrollos en los sistemas y en las concepciones a lo largo de 60 años de investigación. En el segundo capítulo abordaremos la complejidad de la traducción automática. Esta complejidad consiste tanto en las limitaciones de los ordenadores, como en la complejidad de las lenguas naturales. La ambigüedad la hemos considerado un problema clave y hemos prestado atención especial a este fenómeno. En el tercer capítulo explicaremos la metodología que hemos usado para investigar el alcance de la traducción automática. Para determinar la calidad de la traducción automática, las opiniones de los expertos juegan un papel primordial. En el cuarto capítulo presentaremos los resultados de la investigación. Prestaremos atención especial a las ambigüedades en las traducciones automáticas y a las opiniones de los expertos acerca de este fenómeno. A través de este análisis pretendemos ofrecer una imagen del alcance de la TA del par de lenguas español-neerlandés.

Capítulo 1

Introducción a la traducción automática

Este capítulo tiene como propósito ofrecer al lector una imagen introductoria y global de la traducción automática (TA). Primero discutiremos el valor de la TA y a continuación explicaremos el fenómeno de la ‘traducción automática’ en sí y describiremos los diferentes sistemas. Aunque no queremos tratar en detalle las técnicas de estos diferentes sistemas, consideramos esencial saber algo de estas técnicas para poder formar una imagen general de la TA. Concluiremos este capítulo poniendo estas arquitecturas de TA en su contexto histórico, describiendo los 60 años de investigación en TA y el desarrollo de los sistemas.

1.1 El valor de la traducción automática

Antes que nada debemos contestar a la pregunta: ¿por qué queremos automatizar el proceso de traducción? Para poder responder a esta pregunta, primero hay que identificar tres tipos de uso de la TA. Primero Hutchins (2007) menciona la necesidad tradicional de traducciones de calidad “publicable”, en especial la producción de documentación multilingüe para grandes empresas. Sin embargo, muchas veces uno no necesita una traducción perfectamente correcta, sino una traducción que se pueda producir rápidamente (a veces inmediatamente) y que transmita la esencia del texto original, aunque ello suponga valernos de frases gramaticalmente imperfectas, con un léxico extraño y con un estilo rudimentario. Llamamos a este segundo tipo de uso “TA para asimilación”.

Recientemente se ha identificado un tercer tipo de uso relacionado con los intercambios sociales (como correo electrónico, *chats*, etcétera). Este tipo de comunicación tampoco requiere una traducción de alta calidad. Llamamos este tipo de uso "TA para comunicación". En este estudio estamos particularmente interesados en la TA como recurso del traductor. Cada uso específico tiene sus propios requisitos, no obstante, los motivos para automatizar este proceso valen para todos.

A continuación vamos a enumerar las razones que motivan el desarrollo de la TA según Krauwer (2003).

Motivos económicos

Reducir los gastos de traducción es evidentemente un motivo importante para el desarrollo de la TA. Traducir es una obra humana y por lo tanto cara. Un manual de una impresora puede constar de más de 300 páginas. Si el fabricante quiere vender sus productos en todo el mundo con un manual en la lengua de cada región lingüística, hablamos de un factor de alto costo. En el caso de máquinas más complejas, como tanques o aviones, la cantidad de documentos por ejemplar es mucho mayor. Queda claro que hablamos de un montante económico enorme en el que incluso la reducción mínima de algunas traducciones puede proporcionar beneficios significativos.

Velocidad y volumen

La velocidad y el volumen están relacionados con los motivos económicos, pero también tienen sus consecuencias prácticas. El número de pares de lenguas en el que se necesitan traducciones, y la cantidad y tipos de documentos que deben ser traducidos, crecen continuamente. Obviamente un aumento en la velocidad del proceso de traducción conllevará muchas ventajas. Un traductor necesita una a dos horas, dependiendo del tipo de texto, para traducir una página. Aparte de la traducción en sí, el traductor se dedica a leer información de fondo y a descubrir (y a veces inventar) términos técnicos adecuados. Por lo tanto, la traducción de un manual de una impresora puede durar 3 meses, lo cual es demasiado lento para ciertos sectores. Además, es posible encontrar muchos ejemplos de textos que, a pesar de su tamaño, deben ser traducidos a corto plazo, por ejemplo, informes médicos o jurídicos que se necesitan inmediatamente para la tramitación de un caso. Además, la llegada de Internet ha creado una demanda que supera la capacidad de los profesionales de la traducción.

Motivos ideológicos

Una razón importante a favor de la TA es de carácter ideológico, a saber, eliminar las barreras lingüísticas. Quien vive en una región lingüística grande, por ejemplo en los Estados Unidos, tiene fácilmente acceso a grandes cantidades de información, mientras que quienes viven en una pequeña región, por ejemplo en los Países Bajos o Grecia, sólo tienen acceso a la información en su propio idioma o a la información de que han aprendido el idioma. Además, la información en la lengua propia no es accesible para las personas fuera de esa comunidad lingüística. El derecho de tener un acceso igual a toda información disponible, sin tener en cuenta el idioma o la educación, en otras palabras, la eliminación de las barreras lingüísticas, es un motivo importante para el desarrollo de la TA.

Curiosidad

La curiosidad científica es un motivo importante para la investigación en TA. La traducción humana es un proceso muy complejo y difícil de entender y, por ende, una empresa fascinante para los investigadores que estudian cómo se ejecuta con un ordenador este proceso (o cualquier otro proceso que dé el mismo resultado).

1.2 ¿Qué es la traducción automática?

En este apartado abordaremos la traducción automática en sí y describiremos sus diferentes sistemas. Vista la complejidad de los sistemas, no discutiremos este tema en profundidad, sino que explicaremos los diferentes sistemas de forma global. Consideramos necesario ofrecer alguna información sobre la TA, para luego sacar conclusiones sobre las traducciones automáticas y sus principales errores.

Conocemos el fenómeno de ‘traducción automática’ también por su denominación en inglés (*machine translation*). Aunque el término inglés supone que se trata de una máquina “mágica” que convierte un texto de una lengua en otro de una lengua diferente, la realidad es muy distinta. Berner (2003) presenta la siguiente definición de traducción automática:

Machine translation (MT) is the use of computer software to translate text or speech from one natural language into another. Like translation done by humans, MT does not simply involve substituting words in one language for another, but the application of complex linguistic knowledge: morphology, syntax, semantics, and understanding of concepts such as ambiguity

Con los años se han desarrollado diferentes arquitecturas de TA con diversos resultados. Podemos distinguir dos enfoques principales, el enfoque basado en normas y el enfoque basado en un corpus.

1.2.1 Enfoques basados en normas

Las estrategias basadas en normas, de acuerdo con Hutchins (2003), las podemos dividir en tres enfoques tradicionales, a saber: el sistema de traducción directa; el sistema interlingual; y el sistema de transferencia (mostrados en la figura 1).

Sistema de traducción directa

El sistema de traducción directa es el enfoque más sencillo. Está diseñado para un par de lenguas determinado. Se traduce directamente de la lengua fuente (LF) a la lengua meta (LM), su supuesto básico es que el vocabulario y la sintaxis de los textos de la lengua fuente no necesitan ser analizados, sólo lo estrictamente necesario para la resolución de ambigüedades y la correcta

identificación de las expresiones apropiadas en la lengua meta, así como también para la especificación del orden de palabras de la lengua meta. Normalmente, estos sistemas consisten en un único diccionario bilingüe y un programa único para analizar el texto fuente.

Sistema interlingual

El segundo sistema básico es el sistema interlingual, que asume que es posible convertir un texto de LF en representaciones sintácticas y semánticas comunes para más de una lengua. El texto en la LF se transforma en un lenguaje intermedio mediante el componente “análisis”. El texto en la LM se obtiene a partir de la representación del texto en el lenguaje intermedio, mediante el componente “generación”. La estructura del lenguaje intermedio, llamado “interlingua”, es independiente de la de la lengua fuente y de la de la lengua meta y basada en una lengua artificial como por ejemplo el esperanto. Un argumento en su favor es el efecto de economía, con un sólo sistema se puede traducir a varias lenguas, aunque por otro lado, la construcción de tal *interlingua* es un trabajo muy complejo.

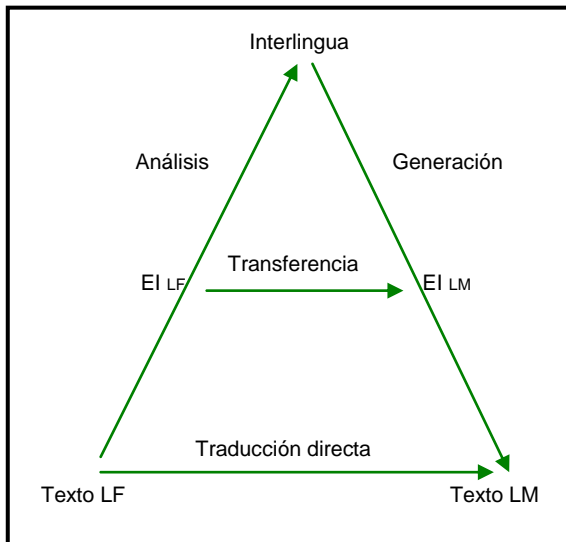


Figura 1.1 El diagrama de la pirámide (Hutchins 2003)

Sistema de transferencia

La tercera estrategia básica, el menos ambicioso “sistema de transferencia”, consiste en tres etapas. La primera etapa (análisis) consiste en la conversión de los textos de la LF en representaciones de transferencia de la misma LF: la Estructura Interfaz de la LF (EI LF). El segundo paso (transferencia) será convertirlas en representaciones de la LM: la Estructura Interfaz de la LM (EI LM). El último paso (generación) consiste en convertir la estructura interfaz de la LM en un texto final de la LM. Mientras el sistema de interlingua requiere una resolución completa de todas las ambigüedades en el texto fuente, de modo que la traducción a cualquier idioma sea posible, el sistema de transferencia sólo aborda las ambigüedades inherentes a la lengua fuente. Los problemas respecto a las diferencias léxicas entre las lenguas se abordan en la segunda etapa (transferencia).

Una manera sencilla para poder entender las relaciones entre los tres sistemas, es reconocer que el sistema de interlingua elimina el componente “transferencia” y que la traducción directa traduce sin el componente “análisis” y “generación” (la traducción sólo existe a partir del componente “transferencia”).

1.2.2 Enfoques basados en corpus

Los enfoques basados en corpus, también llamados “enfoques empíricos”, los podemos distinguir en dos sistemas, a saber, la TA basada en ejemplos y la TA estadística. En este apartado nos basamos en Somers (2003).

Traducción automática basada en ejemplos

La idea básica de la traducción automática basada en ejemplos (TABE) es *reutilizar* muestras reales con sus respectivas traducciones como base de una nueva traducción. A este respecto es comparable (y a veces confundido) con una herramienta usada por los traductores conocida como “memoria de traducción” (MT). Tanto la TABE como la MT se caracterizan por encontrar traducciones correspondientes a una base de datos de traducciones reales. La diferencia entre estos sistemas se encuentra en el proceso de evaluación de las traducciones propuestas y la elección de la solución más adecuada. Mientras en el caso de la TABE este proceso es automático, en el caso de la MT el traductor opta por la solución más apropiada para el texto fuente, en otras palabras, es un proceso humano. El proceso de la TABE consiste en tres etapas: encontrar correspondencias (aquello que la TABE y la MT tienen en común); alinear; y recombinar. En la primera etapa el sistema encontrará, mediante correspondencias con la entrada, muestras de traducciones que pueden contribuir a la traducción. No profundizaremos en los detalles técnicos de esta etapa. La segunda etapa, la alineación, consiste en identificar las partes útiles de la traducción correspondiente. En el caso de la MT los traductores humanos realizan esta etapa de “alineación”, mientras que en el caso de la TABE ésta es automatizada. Si, por ejemplo, queremos traducir la frase 1a a partir de los ejemplos 1b y 1c, debemos determinar qué partes de los equivalentes holandeses se corresponden con las partes en verde en los ejemplos.

1. a. Compra un libro sobre la política internacional
- b. **Compra un** ordenador. *Hij koopt een computer.*
- c. Leí **un libro sobre la política internacional**. *Ik las een boek over internationale politiek.*

El tercer paso, llamado “recombinación”, *recombina* las partes correspondientes. Cuando sabemos qué partes de los ejemplos reutilizamos, tenemos que intentar que las partes correspondan de manera legítima. En el ejemplo 2 (procedente de Somers (2003)) daremos un ejemplo que muestra la complejidad de este proceso. Se trata de una frase en español traducida al alemán, una lengua que hace una diferencia marcada entre objeto y sujeto. La frase alemana que corresponde

con la frase “el chico guapo” difiere en los ejemplo 2b y 2c. El artículo correcto depende de la función gramatical que desempeñe “el chico” (sujeto o objeto). Para determinar cuál de las frases es adecuada se debe tener nociones de gramática del alemán.

2. a. El chico guapo entró el cuarto.
- b. El chico guapo comió el desayuno. *Der schöne Junge aß seinen Frühstück.*
- c. Vi el chico guapo. *Ich sah den schönen Jungen.*

Traducción automática estadística

En su forma auténtica, la traducción automática estadística no usa datos lingüísticos tradicionales. La esencia de este método es alinear frases, grupos de palabras y palabras individuales de textos paralelos y calcular las *probabilidades* de que una palabra en una frase de una lengua se corresponda con una palabra en una frase de una traducción con la que está alineada. Dado que la TA estadística genera sus traducciones a partir de métodos estadísticos basados en corpus de textos bilingües, la disponibilidad de un corpus grande de traducciones fiables es una característica esencial de este sistema. Se suele ver este método como “anti-lingüístico”. La idea de este sistema es modelar el proceso de traducción en términos de probabilidades estadísticas. Por ejemplo, si queremos traducir la frase 3a, encontramos las frases 3b y 3c entre las traducciones posibles:

3. a. Compra un libro sobre la política internacional
- b. *Hij koopt een boek over internationale politiek.*
- c. *Vanochtend heb ik mijn tanden gepoetst.*

Es probable que la frase 3b sea una buena traducción, pero es menos probable que la frase 3c lo sea también. Por cada par de frases F (fuente) y M (meta) hay una probabilidad (P): $P(M|F)$. El proceso de traducción es cuestión de encontrar el mayor valor por $P(M|F)$. La probabilidad de que una palabra en el texto fuente represente una palabra en el texto meta, se calcula sobre la base de un corpus alineado. Es un cálculo laborioso, pero que se hace una vez para un corpus paralelo, teniendo en cuenta la distribución relativa de todas las palabras en un corpus. Por ejemplo, de un corpus puede resultar que la probabilidad es de 0,610 que se traduzca "een" como "un", y de 0,170 que se traduzca como "una". Lo llamativo de este método, es la ausencia absoluta de un conocimiento lingüístico en el proceso. Si el sistema relaciona “la” con “flor”, no es porque “sepa” algo de la concordancia de género, sino porque esta combinación es la más frecuente.

1.2.3 Conclusión

Cuando se publicaron los primeros informes acerca de la traducción automática estadística, los investigadores estaban sorprendidos por los resultados aceptables de este método. Sin embargo, cuando progresaba la investigación, se dieron cuenta de que las posibilidades de mejorar el rendimiento de estos sistemas eran limitadas. Muchos de los errores se podían corregir con un conocimiento lingüístico mínimo. Ni la traducción automática basada en ejemplos, ni la basada en estadística, han resultado mejor que los sistemas basados en reglas, aunque cada método ha sido prometedor en ciertos terrenos. Eso favoreció el desarrollo de “sistemas híbridos”, que reconocen las ventajas y desventajas específicas de cada método.

1.3 La evolución de la traducción automática e investigación

En este apartado describiremos la evolución y el desarrollo de los distintos sistemas de TA (véase el apartado 1.2) en su contexto histórico. Además, discutiremos el cambio en las concepciones de los investigadores de TA. En este apartado nos basamos en Hutchins (2003 y 2007).

Precursores y pioneros, 1933-1956

Aunque en el siglo XVII ya se pueden encontrar ideas de lenguas universales y diccionarios “mecánicos”, no fue hasta el siglo XX cuando se hicieron las primeras sugerencias prácticas para una traducción automática. En el año 1933 se otorgaron dos patentes en Francia y Rusia a, respectivamente, George Artsrouni y Petr Trojanski. De las dos patentes, la de Trojanski era la más significativa, debido a que no sólo propone un método para un diccionario bilingüe automático, sino también un esquema para codificar funciones gramaticales interlinguales (basado en el esperanto) y un diseño de cómo podría funcionar los componentes análisis y síntesis. Andrew Booth y Warren Weaver no sabían nada acerca de ninguno de estos precursores cuando se conocieron en 1946 y presentaron las primeras ideas para usar los recién inventados ordenadores para la traducción de lenguas naturales. Para entonces, la idea de traducción mecánica había surgido de manera independiente y en el año 1949 Warren Weaver (el director de la Fundación Rockefeller) propuso ideas específicas para solucionar los problemas evidentes de ambigüedad, basadas en su conocimiento de criptografía, estadística, lógica y lenguajes universales. Este memorando fue el mayor estímulo de investigación en la TA en los Estados Unidos. Un par de años después la investigación en TA empezaba en varias universidades estadounidenses y en 1954 tuvo lugar la primera demostración pública (una colaboración entre IBM y la Universidad de Georgetown). Aunque se usaba un vocabulario y una gramática restringidos, fue lo suficientemente impresionante para estimular la financiación de la TA en los Estados Unidos y para inspirar proyectos de TA en todo el mundo. Ya quedó claro que la automatización total de una traducción de alta calidad (*a fully automatic high quality translation (FAHQT)*) era prácticamente imposible, y que la intervención humana, fuera antes o sea después de los procesos informáticos, era esencial. Algunos esperaban que sería sólo

una medida provisional, pero la mayoría esperaba que la intervención humana siempre sería necesaria.

La década del optimismo, 1954-1966

Cuando la investigación en TA comenzaba, los métodos de investigación tendían a polarizarse en, por un lado, los enfoques empíricos que a menudo adoptaban métodos estadísticos para “descubrir” regularidades gramaticales y léxicas, y, por otro lado, los enfoques teóricos con su base en la investigación lingüística, lo que trajo consigo el comienzo de la investigación en lo que más tarde se llamaría “lingüística computacional”. En esta época se solían describir estos métodos contrastivos respectivamente como “fuerza bruta” y “perfeccionista”. En esta primera década se pueden ver los inicios de los tres enfoques básicos para la TA (véase el apartado 1.2.1. *enfoques basados en normas*). El primer enfoque fue el método de “traducción directa”, donde las normas de programación se habían desarrollado específicamente para la traducción de una lengua fuente a una lengua meta en particular, con un análisis y reorganización sintáctica mínima. El segundo enfoque fue el método interlingual, basado en la representación independiente de la lengua. Este tipo de traducción consiste en un trabajo programado en dos etapas, de lengua fuente a interlingua y de interlingua a lengua meta. El tercer enfoque es el método menos ambicioso de transferencia, en el que la representación del texto fuente se traduce mediante la etapa de transferencia a su representación equivalente del texto meta. En la mayoría de los casos, los “empiristas” usaban el método de traducción directa, frecuentemente usando análisis estadísticos de textos reales para derivar las reglas del diccionario, a menudo de carácter *ad hoc*, con muy poca fundamentación teórica. Los perfeccionistas se basaban explícitamente en la teoría, prestando atención especial a los métodos de análisis sintácticos. Algunos grupos de investigadores perseguían el ideal de “interlingua”, y consideraban que sólo la investigación del proceso de pensamiento humano (lo que más tarde se llamó inteligencia artificial) podría solucionar los problemas de TA. Las personas más pragmáticas entre ellos se concentraron en los modelos de transferencia, posponiendo los problemas semánticos a una etapa posterior.

En esta época, por motivos políticos y militares, casi toda la investigación estadounidense se centraba en la traducción de ruso a inglés. Además, tenemos que tener en mente que los ordenadores de entonces no tenían la misma capacidad que los de hoy. Esta época se caracterizaba por el optimismo y las altas expectativas. El desarrollo de la informática y de la lingüística formal, en particular en el ámbito sintáctico, eran prometedores. Había muchas predicciones de avances inminentes y sistemas completamente automáticos en pocos años. Sin embargo, la desilusión creció cuando la complejidad de los problemas lingüísticos eran más evidentes, los investigadores se enfrentaron con barreras semánticas sin soluciones aparentes.

El informe ALPAC y sus consecuencias

En 1964, el gobierno estadounidense, que patrocinó la investigación, estaba cada vez más preocupado por la falta de progreso, por esa razón formaron el *Automatic Language Processing Advisory Committee* (ALPAC), que investigó el estado de la TA y sus futuras perspectivas. El ALPAC concluyó en su famoso informe que la TA era más lenta, menos precisa y dos veces más cara que la

traducción humana y que no había perspectivas inmediatas o previsibles de un método de TA útil (ALPAC 1966). No se veía ninguna necesidad de invertir más en la investigación en TA, en su lugar se recomendaba el desarrollo de ayudas para el traductor como por ejemplo diccionarios automáticos. Paradójicamente, el ALPAC rechazaba la TA porque requería *post edición* (a pesar de que también se revisan las traducciones humanas antes de su publicación) y porque se suponía que los usuarios exigían traducciones de alta calidad, a pesar de que los organismos patrocinadores estaban principalmente interesados en la recopilación de información y análisis, en lo que una menor calidad era aceptable. Aunque muchos condenaban el informe como “corto de vista”, la influencia del ALPAC fue significativa; durante muchos años se les denegaron todos los fondos oficiales para investigaciones dedicadas a este fin en los Estados Unidos y tuvo gran impacto en otras partes del mundo como en la Unión Soviética y Europa.

La década silenciosa, 1967-1976

No obstante, la investigación no se detuvo por completo. Incluso en los Estados Unidos continuó durante algunos años más, pero se produjo un cambio de dirección. Mientras que los métodos de traducción directa dominaban en la “primera generación” (el período anterior a la ALPAC), en la “segunda generación” (posterior a la ALPAC) dominaban los métodos indirectos, tanto los métodos interlinguales como los de transferencia. En los Estados Unidos la actividad principal se concentró en traducciones de inglés a ruso de materiales científicos y técnicos. En Canadá y Europa las necesidades eran muy distintas. En Canadá, la política bicultural producía una demanda de traducciones de inglés a francés más allá de la capacidad de la profesión de traductor. Los problemas de traducción en la Comunidad Europea eran igualmente urgentes, con una creciente demanda de traducciones de documentaciones científicas, técnicas, administrativas y jurídicas desde y hacia todos los idiomas comunitarios. En 1976, el sistema Systran se instaló en la Comisión de la Comunidad Europea para la traducción de la documentación rápidamente creciente. En el mismo año, otro sistema exitoso apareció en Canadá, el sistema de Meteo para la traducción de boletines meteorológicos desarrollado por la universidad de Montreal. A mediados de los años 70, el futuro del método interlingual estaba en duda. Los problemas principales de este método se atribuyeron a la rigidez de los niveles de análisis y, en particular, a la pérdida de información del texto fuente. Como consecuencia en este momento, a muchos les parecía que el método de transferencia ofrecía mejores perspectivas.

Sistemas comerciales y la reactivación de la investigación, 1976-1989

En los años 80 tuvo lugar el desarrollo de una gran variedad de sistemas de TA en cada vez más países. Una serie de estos sistemas centrales continúa usándose hasta el día de hoy. Aparte de Systran, que hoy en día ofrece muchos pares de idiomas, estaba Logos (que servía de alemán a inglés y de inglés a francés); los sistemas desarrollados en la Organización Panamericana de la Salud (español a inglés e inglés a español); el sistema Metal (alemán a inglés); y los sistemas para la traducción de inglés a japonés y de japonés a inglés de empresas informáticas japonesas. La gran presencia de ordenadores y de procesadores de texto creaba un mercado para sistemas de TA más

baratos, explotado por empresas norteamericanas y europeas como ALPS, Weidner, Linguistic Products y Globalink y por muchas empresas japonesas como Sharp, NEC, Oki, Mitsubishi y Sanyo. Otros sistemas informáticos aparecieron en China, Taiwán, Corea, Europa del Este, la Unión Soviética, etcétera.

A lo largo de los años 80 se continuó con la investigación de métodos y técnicas más avanzados. La estrategia dominante en esta década era la traducción “indirecta” a través de representaciones de intermediarios, a veces de carácter interlingual, usando tanto análisis semánticos como morfológicos y a veces usando componentes no lingüísticos como una “base de conocimiento”. Los proyectos más notables de este período fueron el proyecto de Eurotra (Grenoble), que tenía como objetivo construir un sistema de transferencia multilingüe para la traducción entre todas las lenguas comunitarias; el proyecto de Mu (Kyoto), que ofrece una gran variedad de métodos tanto de transferencia como interlinguales; el proyecto de DLT (Utrecht) de la compañía de software BSO, que desarrolló un sistema interactivo multilingüe basado en redes informáticas. El proyecto realizó un paso significativo en la construcción de una base de datos léxica y en sus últimos años propuso la construcción de un Banco de Conocimiento Bilingüe de un corpus de traducciones (humanas). En este sentido, podemos decir que estaba anticipándose a lo que más tarde serán los sistemas basados en ejemplos. Otro proyecto destacado era el de Rosetta (Eindhoven) en Philips. Una característica importante de este proyecto era la reversibilidad de las gramáticas, es decir, el conjunto de reglas gramaticales funcionaba en una dirección para el análisis sintáctico y semántico de una lengua y en la otra dirección para la generación de frases correctas de esta lengua. La reversibilidad ha vuelto a ser un tema importante en muchos proyectos de TA posteriores. La Universidad Carnegie-Mellon (Pittsburgh) avanzó en los proyectos de sistemas basados en el conocimiento. La base del sistema no consiste exclusivamente en datos morfológicos, gramaticales y traducciones equivalentes, sino también en información no lingüística y conceptual.

El gran giro

A principios de los años 90 hubo un gran giro en la evolución de la TA. En primer lugar, un grupo de IBM publicó los resultados de los experimentos de un sistema (Candide) basado en métodos exclusivamente métodos estadísticos. Los investigadores, enfocados en sistemas basados en lingüística, estaban sorprendidos de los resultados aceptables de estos experimentos.

En segundo lugar, algunos grupos japoneses empezaron a usar métodos basados en corpus de traducciones, es decir, usando el enfoque que ahora se llama “traducción basada en ejemplos”. En ambos métodos la característica distintiva era la ausencia de reglas sintácticas o semánticas en el análisis de los textos y en la selección de equivalentes léxicos. Ambos métodos diferían de los métodos basados en reglas. Vemos este período como la vuelta del “empirismo” de la primera década y el desafío al “racionalismo” de los años 60 y 80. Aunque hoy en día la investigación de TA se enmarca principalmente en el campo de la TA estadística, los dos métodos basados en corpus convergirán en el sentido de que la TA estadística usa la alineación de frases y de datos lingüísticos, mientras que la TA basada en ejemplos usa técnicas estadísticas. Por lo tanto, resulta cada vez más difícil distinguir las características de los dos modelos. Aunque la innovación principal

desde los años 90 ha sido el crecimiento de los métodos basados en corpus, la investigación de métodos basados en reglas, tanto de sistemas de transferencia como interlinguales, continuó. Gracias a la expansión de las metodologías en la última década los investigadores se han dado cuenta de que ningún método produce una TA de buena calidad, y que los futuros modelos serán “híbridos” combinando los mejores elementos de los métodos basados en reglas, con los basados en ejemplos y los estadísticos.

Otra característica de los años 90 fue el cambio de foco de la investigación en TA. Cambió de mera investigación a aplicaciones prácticas, como el desarrollo de la estación de trabajo del traductor (en inglés: *translators workstation*), el lenguaje controlado, los sistemas con dominios restringidos y la aplicación de componentes traductores en los sistemas de información multilingüe.¹

La TA en Internet

Concluimos este apartado con las aplicaciones de la TA en Internet. No queremos decir que el desarrollo de la TA termine con la TA en Internet, sólo queremos subrayar la importancia de este fenómeno en la actualidad. Desde mediados de los años 90 Internet ha sido una influencia poderosa en el desarrollo de la TA. La llegada de Internet ha producido una demanda de traducciones inmediatas en línea, algo que los traductores imposiblemente puedan cubrir. En Internet toda la información es fácilmente accesible. Gracias a las técnicas avanzadas sabemos inmediatamente qué pasa al otro lado del mundo. Una de las pocas barreras que nos impide el acceso a esta información es el idioma, por lo tanto hay una alta demanda a traducciones inmediatas de sitios web. La primera aplicación de TA gratuita en línea fue *Babelfish*, un sistema basado en reglas, localizado en el sitio web de *Altavista* en el año 1997. Esta aplicación ofrece una versión de Systran para traducir del francés, del alemán y del español al inglés y a la inversa, y más tarde muchos otros pares de idiomas (pero todavía no ofrece el par de lenguas español – neerlandés). Su objetivo es eliminar las barreras lingüísticas en la web (Somers 2003: 191). En 2006 el *Traductor Google* llegó al mercado. El *Traductor Google*, desarrollado por Franz-Josef Och (ahora jefe del departamento de TA en *Google*), se basa en métodos estadísticos. La aplicación de *Google* no sólo se distingue por su enfoque estadístico, sino también por el gran número de pares de lenguas que ofrece (entre otros español-neerlandés). A menudo la calidad de las traducciones de estas aplicaciones en Internet es mala debido a la naturaleza coloquial de muchos textos fuente; el lenguaje en Internet es muchas veces incoherente, lleno de acrónimos y abreviaturas, juegos de palabras, chistes, etcétera. Sin embargo, hay una gran demanda de estas traducciones “crudas”.

1.4 Conclusión

En este capítulo hemos intentado dar una imagen introductoria de la TA. Hemos explicado los diferentes sistemas que se usan hoy en día. Tanto los sistemas basados en normas como los basados en corpus tienen sus ventajas, pero los expertos esperan que los sistemas híbridos serán los

¹ No tomamos en consideración en esta tesina la tercera innovación la “traducción de voz”.

más prometedores. En sesenta años de investigación las concepciones de los investigadores han fluctuado mucho, desde altas expectativas hasta grandes desilusiones. La demanda de traducciones inmediatas en línea ha creado una nueva necesidad de TA. En un mundo tan abierto y avanzado muchas veces la única barrera para tener acceso a cierta información es la lengua. En el mundo del internet la información es muy dinámica, y por tanto hay una alta demanda de traducción inmediata. En este aspecto el valor de la TA en línea es de gran importancia. En esta tesina nos concentramos en la TA en Internet, para ser más específicos, en el traductor automático *Google*.

En el segundo capítulo discutiremos los problemas que surgen cuando se traduce con la ayuda de un traductor automático.

Capítulo 2

La complejidad de la traducción automática

En este capítulo intentaremos contestar a la pregunta: ¿por qué traducir es tan difícil para los ordenadores? A menudo la investigación en la TA ignora los avances de la teoría lingüística. La investigación de la TA ha sido considerada como ‘trabajo de ingenieros’ pero nosotros opinamos que la investigación lingüística es de suma importancia. En la investigación lingüística, el centro de atención siempre ha sido los fenómenos lingüísticos relacionados con la lengua inglesa. En otras palabras, la teoría lingüística no siempre ha investigado la lingüística contrastiva, o sea, la manera en la que las diferentes lenguas utilizan distintas medidas para expresar determinados significados. La falta de atención a los otros idiomas es una de las razones por las que la teoría lingüística no tiene un impacto marcado en la investigación de la TA. En este capítulo no abordamos las complejidades técnicas de la TA, sino que enmarcamos este asunto en el campo lingüístico. Como muchas complejidades lingüísticas están relacionadas con un determinado par de lenguas, intentamos concentrarnos en los fenómenos lingüísticos que surgen en el par de lenguas español y neerlandés.

2.1 Limitaciones y problemas

2.1.1 Lenguas naturales

La complejidad de la TA está relacionada con la riqueza de las lenguas naturales. El procesamiento de lenguas naturales es algo muy complejo que todavía no entendemos completamente. ¿Qué es lo que hace tan difícil procesar estas lenguas naturales? Cuetara Priede (2007:10) presenta la siguiente definición de las lenguas naturales:

Una lengua natural es un conjunto de símbolos vocales arbitrarios por medio del cual se comunican los seres humanos. Tal conjunto está organizado en un sistema de relaciones, sintagmáticas y paradigmáticas, muy cohesionadas que hace posible que el habla sea un sistema de comunicación muy preciso y especializado.

Esta arbitrariedad de la lengua natural también la menciona el lingüista Charles F. Hockett (1960) en sus quince rasgos definitorios de la lengua natural. Algunos de los rasgos definitorios que discute Hockett también se encuentran en la comunicación animal y en las lenguas formales, pero sólo las lenguas naturales tienen todos estos rasgos. Mencionaremos aquellos que muestran la complejidad de una lengua natural a la hora de automatizarla.

Con la arbitrariedad Hockett quiere decir que no existe correlación entre el signo (*signifiant*) y el significado (*signifié*). Por ejemplo, los fonemas que crean la palabra “libro” no tienen ninguna relación con el concepto de “libro”. No hay ninguna razón para que el concepto de un conjunto de hojas debiera ser un “libro” y no cualquier otra combinación de fonemas². Otro rasgo interesante de Hockett es la productividad de una lengua, es decir que las reglas de la gramática permiten la creación de una infinita cantidad de construcciones sintácticas; en otras palabras, lo que Chomsky (1977) llama “gramática generativa”, permite la creación de oraciones que jamás han sido creadas, pero que pueden ser entendidas. Por ejemplo, la frase <<Con el perro en mi armario azul hablo sobre mis vacaciones en Marte durante el verano de 1970>>, probablemente es una frase que nunca antes ha sido creada, pero que un hispanohablante puede entender perfectamente.

Por último queremos mencionar el rasgo de transmisión cultural, que quiere decir que la lengua humana no es innato, sino un producto de la evolución histórica que se transmite entre generaciones. Por consiguiente, una lengua natural está vinculada con la evolución cultural e histórica de una región. Es obvio que por lo menos estos tres rasgos producen problemas a la hora de automatizar una lengua.

2.1.2 Limitaciones

Una razón que explica la dificultad que tienen los ordenadores con la traducción es simplemente el hecho de que traducir en sí es difícil, incluso para los traductores humanos. Un traductor se dedica a

² Aquí no mencionamos las onomatopeyas, que sí muestran una relación entre el signo y el significado (aunque los perros españoles dicen “guau” y los perros holandeses dicen “waf”).

producir un texto “equivalente” en una lengua fuente, que tenga que servir autónomamente, siendo fiel al texto fuente y comprensible en la lengua meta³. El término “equivalencia” es difícil de precisar. Tenemos que tener en cuenta que las lenguas difieren, y, por consiguiente, que una lengua meta no siempre permite expresar el mismo contenido que una lengua fuente. Por ejemplo, “nosotras” en español es una palabra más específica acerca del género que el “wij” en neerlandés. Crear un texto equivalente en la lengua meta no es un trabajo sencillo. Además, no tenemos que olvidar que traducir es un trabajo creativo. El traductor no sólo debe “inventar” traducciones para los nuevos términos que aparecen en el texto fuente, sino que también debe actuar como intermediario entre dos culturas. Muy a menudo el traductor está en un conflicto en el que tiene que optar, por un lado, entre “modernizar” o “historizar”, y por otro entre “naturalizar” o “exotizar” (Holmes 2004). Por ejemplo, ¿cómo se traduce el pájaro tropical latinoamericano “agapornis” al neerlandés? ¿Se reemplaza por un pájaro equivalente holandés o se deja como un nombre éxotico para los lectores holandeses? Queda claro que incluso para los traductores humanos el trabajo de traducir es muy difícil y requiere mucha creatividad. Los ordenadores son fundamentalmente dispositivos que siguen reglas. Seguir las reglas puede producir cierto tipo de creatividad, pero no el tipo de creatividad que requiere la traducción de un texto (Somers 2003: 120). Encontrar nueva terminología, un neologismo, es más bien una cuestión de inventar una regla que seguir una regla. La mediación intercultural es un trabajo muy complejo: el traductor no sólo tiene que ser capaz de extraer el sentido de un texto, sino también tiene que ser capaz de prever el sentido que extraiga el lector potencial del texto meta.

En el presente estudio partimos de la idea de que los lectores del texto fuente y meta comparten la misma cultura y el mismo conocimiento de fondo. Además, ignoramos los problemas con respecto a la nueva terminología. En otras palabras, nos concentramos en la traducción de textos escritos de una lengua a otra, transmitiendo el mismo contenido aproximado, sin problemas debidos a la nueva terminología y dirigiéndonos a los lectores que comparten la misma cultura que los lectores del texto fuente.

Aparte de que traducir en sí es difícil, los ordenadores tienen que hacer las traducciones dentro de sus limitaciones. Arnold (en Somers 2003: 121) menciona cuatro limitaciones de los ordenadores que están en la raíz de los problemas de la TA, a saber, la incapacidad de los ordenadores de:

1. *Realizar tareas no específicas:*

Para poder traducir de una lengua a otra, un ordenador necesita reglas precisas. Gran parte de la dificultad del procesamiento de la lengua natural en general, es la dificultad de encontrar reglas suficientemente precisas de ideas intuitivamente muy claras, como por ejemplo: “generalmente, el sujeto precede al verbo en neerlandés”. El problema aquí es la palabra “generalmente”, que desgraciadamente aparece en muchas reglas. Un ordenador necesita reglas precisas, lo que no siempre es posible.

2. *Aprender cosas nuevas:*

³ En el presente estudio nos limitamos a la complejidad concerniente a la traducción del contenido. No tomamos en consideración aspectos con respecto al estilo.

El mundo y la lengua están cambiando continuamente, por lo tanto es muy importante para un traductor estar bien enterado de muchos ámbitos, lo que significa: aprender cosas nuevas. Y aprenderlas requiere creatividad, es decir, el traductor debe inventar las reglas en lugar de seguir las reglas. Un ordenador sólo sabe lo que se le ha dicho en forma de reglas explícitas, escritas por humanos.

3. Realizar “razonamientos de sentido común”:

La tercera limitación es que los ordenadores no pueden realizar “razonamientos de sentido común”. Hay varias razones que explican esta limitación, quizá la más seria sea el hecho de que un razonamiento de sentido común implica, literalmente, millones de hechos en el mundo. Por ejemplo, el agua es húmeda, el fuego es caliente, los hombres no pueden estar embarazados, la mayoría de las personas tiene dos pies, un caballo es más grande que un ratón. La tarea de codificar esta gran cantidad de conocimiento es de proporciones enormes, y fuera de la capacidad de los ordenadores modernos.

4. Enfrentarse a problemas con un gran número de soluciones posibles

La cuarta limitación tiene que ver con la incapacidad de los ordenadores de enfrentarse a un gran número de soluciones posibles. Eso quiere decir que, aunque los problemas hayan sido especificados precisamente, el traductor automático debe elegir entre muchas soluciones, a veces entre un número de soluciones infinito.

Las limitaciones de los ordenadores también pueden tener sus ventajas. Por ejemplo, un texto traducido automáticamente cuenta con la garantía de que no repite u omite partes del texto fuente, además no es probable que haga errores ortográficos (salvo errores de puntuación o de capitalización). Otra ventaja es que, aunque el traductor automático no siempre puede encontrar la traducción correcta de una frase o palabra, no hace ‘adivinaciones salvajes’, como el traductor humano a veces hace, traduciendo por ejemplo “su volumen rotundo” por “zijn ronde omvang” en lugar de “imposante omvang” (un error encontrado en una traducción humana). Consideramos la consistencia como ventaja principal de la TA, aunque, obviamente, eso no compensa sus desventajas.

2.1.3 Problemas lingüísticos

Hemos explicado las limitaciones que están en la raíz de los problemas de traducir automáticamente. Se puede prever que la complejidad de las lenguas naturales y las limitaciones de ordenadores producen muchos problemas a la hora de traducir automáticamente. Los problemas principales que surgen según Krauwer (2003) son los siguientes:

La ambigüedad

Es uno de los más grandes problemas de los traductores automáticos. Debido a las limitaciones de los ordenadores, el traductor automático se enfrenta a muchas más ambigüedades que el traductor humano. Como este fenómeno es tan complejo, dedicaremos el apartado 2.2. a este asunto.

La robustez

Con *robustez* de un traductor automático nos referimos a la capacidad de los sistemas de reaccionar a las entradas inesperadas. Este tipo de entradas puede consistir en nuevas palabras, reglas gramaticales inesperadas (lo que fácilmente puede ocurrir, dado el hecho de que ninguna lengua tiene un registro gramatical de todos sus fenómenos, excepciones y rasgos dialectales tan exasustivo que podamos decir que está completo). Asimismo, el texto fuente puede consistir en frases gramaticalmente incorrectas, por ejemplo en palabras equivocadas, inflexiones, etcétera.

Las lenguas difieren

Muchas veces las lenguas usan distintas formas para expresar el mismo contenido; sin embargo, hay suficientes similitudes para encontrar las estructuras y palabras correspondientes. No obstante, con esto se corre el riesgo de crear una lengua de “traslacionismo”, en la que se traduce en la estructura de la lengua fuente, o sea, la traducción literal. A menudo, las lenguas utilizan otras formas para expresar cierto contenido, o utilizan estructuras radicalmente diferentes para expresar el mismo contenido. Obviamente existe un sinnúmero de ejemplos en los que el neerlandés utiliza diferentes estructuras, palabras o expresiones que el español, así que en este apartado sólo discutimos las diferencias más aparentes e interesantes para la TA⁴.

No siempre se puede encontrar una palabra equivalente en la misma clase de palabra (ejemplo 4). En neerlandés no se puede traducir “normativas” como un adjetivo en este contexto (*het normatieve gebrek). Otro ejemplo lo encontramos en la frase 5; aquí el español utiliza sustantivos mientras que en neerlandés nos vemos forzados a reemplazarlos por verbos.

- | | |
|--|--|
| 4. Las carencias normativas de Internet. | Het gebrek aan normen op internet. |
| 5. Los enfermeros tienen que estar informados de la aparición de estos síntomas, de su prevención y de su tratamiento . | De patienten moeten weten dat deze verschijnselen zich [kunnen] voordoen , en hoe ze zijn te voorkomen en te behandelen . |

Siempre que se habla desde cierta perspectiva, el contraste entre lenguas puede suponer un cambio de perspectiva. En el ejemplo 6, el verbo español requiere otro sujeto que el verbo neerlandés. Mientras que en la frase española la persona que paga el dinero es el sujeto, en la frase neerlandesa la persona que recibe el dinero es el sujeto.

- | | |
|--|---|
| 6. Sigo haciendo lo mismo, pero ahora me pagan por ello . | Ik doe nog steeds hetzelfde, maar nu krijg ik er geld voor . |
|--|---|

El ejemplo 7 nos muestra que el español puede ser más preciso acerca del género de un grupo de personas que el neerlandés. En neerlandés no podemos mantener este marco de género, sin parecer evasivos (“sommige van deze vrouwen”).

⁴ Los ejemplos son de Linn & Slager

7. **Algunas de nosotras** ya llevamos tiempo trabajando en la empresa

Sommigen van ons werken al een poos bij dit bedrijf.

Otro ejemplo que muestra que en español a veces se utilizan muy diferentes estructuras de las del neerlandés es la frase 8. La estructura que el neerlandés utiliza aquí es totalmente diferente de la española. No sólo el sujeto es explícito en la frase neerlandesa, también necesita la preposición "met".

8. **-¿Cuántos sois? - Somos dos.**

Met z'n hoevelen zijn jullie? We zijn met z'n tweeën.

Además, no tenemos que olvidar que a veces hay más de una traducción correcta para una frase, pero una es más idiomática que la otra. El ejemplo 9 muestra varias soluciones para la traducción de "¿Qué hora es?"

- 9. a. Wat is het uur?
- b. Wat is de tijd?
- c. Hoe laat is het?

Aunque las tres opciones son gramaticalmente correctas, un caso es más idiomático que el otro (en este caso la opción 9c).

2.2 Ambigüedad

De todas las dificultades a las que los traductores automáticos se enfrentan, los problemas con respecto a la ambigüedad son quizás los más significativos. Krauwer (2003) opina que la ambigüedad pertenece a las clases de problemas que no tienen soluciones en la TA. En este apartado primero discutiremos los diferentes tipos de ambigüedad en las lenguas naturales y en la traducción automática y abordaremos tanto las ambigüedades neerlandesas como las españolas. En el capítulo 4, analizaremos las ambigüedades en las traducciones automáticas.

2.2.1 Ambigüedad léxica

La polisemia y la homografía

Según explica Vázquez Ayora (1977) la causa de la ambigüedad léxica es la polisemia. Una palabra, o 'unidad', es polisémica si muestra una variedad de significados, o "acepciones", relacionados de algún modo entre sí. Una unidad es ambigua cuando no sabemos a cuál de las acepciones se refiere, y sólo el contexto puede resolver este problema. En muchos casos el significado de un término está indicado por las construcciones gramaticales en que aparece. Las palabras del vocabulario de uso

habitual son por lo general polisémicas. El desarrollo de una cultura trae consigo el enriquecimiento polisémico de las palabras (de ahí la importancia para el traductor que se dedica a obras de actualidad). En la figura 2.1 presentamos las principales acepciones de la unidad “centro”. Para determinar la acepción adecuada de la unidad “centro”, tenemos que conocer el contexto. Por ejemplo: ¿se refiere a un instituto cultural o a la parte central de una ciudad?

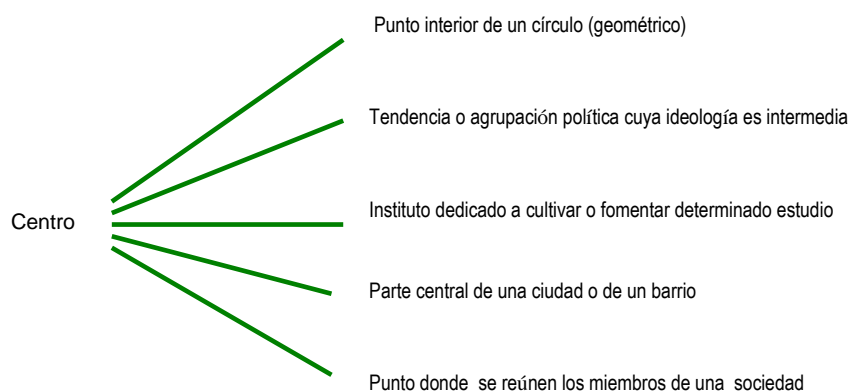


Figura 2.1 La polisemia de la palabra “centro” (fuente: DRAE).

Una unidad es homográfica si se refiere a dos o más significados diferentes. Por ejemplo “banco” puede referirse tanto a un asiento como a un establecimiento financiero. No obstante, en la práctica las diferencias entre las palabras homógrafas y polisémicas son difíciles de tratar.

La ambigüedad categorial

Otra categoría de la ambigüedad léxica es la ambigüedad categorial. Hablamos de ambigüedad categorial cuando una unidad se puede referir a más de una categoría gramatical (por ejemplo sustantivos, verbos, adjetivos, etcétera) dependiendo del contexto. Hay muchos ejemplos de este tipo de ambigüedad en neerlandés, por ejemplo la palabra “licht” puede referirse a un sustantivo (luz), a un adjetivo (ligero) o a un verbo (levantar). Y en español, “vino” puede referirse a un sustantivo o a un verbo. Según explican Colorado y de Borja (2007) los pares más comunes son los de sustantivo y verbo, ya que hay muchos sustantivos que pueden funcionar como verbos (por ejemplo “cuenta”). Un ejemplo extremo, pero no raro, es la palabra neerlandesa “rond” que puede llegar a funcionar como sustantivo (10a), verbo (10b), adjetivo (10c), preposición (10d) y adverbio (10e).

10. a. Nederland heeft de laatste **ronde** van het WK niet gewonnen.
 b. Ik **rond** het bedrag af op 10 euro.
 c. Ik wil een **ronde** tafel kopen.
 d. Wij zitten **rond** de tafel.
 e. Ik kom er **rond** voor uit dat ik van studeren houd.

Otro fenómeno interesante son las palabras compuestas, algo que aparece mucho en neerlandés. Obsérvense las diferencias entre las siguientes palabras compuestas:

- 15. a. Tarwemeel (harina hecha de trigo)
- b. Pannenkoekenmeel (harina para preparar crepes)
- c. Kindermeel (harina para niños)

En neerlandés las tres palabras están compuestas de la misma manera, aunque la relación entre las palabras es muy distinta.

2.2.3 Referencias anafóricas y catafóricas

Podemos considerar las referencias anafóricas y catafóricas como un tipo de ambigüedad. Una anáfora es una referencia indirecta a una entidad mencionada de forma explícita en otro lugar del texto (Somers 2003). La referencia catafórica es un mecanismo simétrico a la referencia anafórica. La diferencia es que en el caso de una referencia catafórica una unidad del texto alude a otra unidad que aparece posteriormente. Como las referencias anafóricas son más frecuentes, sólo abordamos estas en el análisis de las traducciones automáticas. En muchos casos es importante identificar el antecedente, o sea, el objeto al que alude la referencia anafórica, para traducir correctamente. En los ejemplos 16a y 16b la referencia anafórica alude a un antecedente diferente.

- 16. a. De soldaten schoten op de vrouwen en **enkelen van hen vielen neer**.
- b. De soldaten schoten op de vrouwen en **enkelen van hen misten**.

En este caso, la lengua meta (el español) distingue el género de las referencias indirectas, es decir, hace una diferencia entre “algunas de ellas” y “algunos de ellos”. Eso muestra la importancia de la identificación del antecedente a la hora de traducir este tipo de oraciones. Podemos considerar la anáfora como un tipo de ambigüedad en el que el antecedente no se conoce con certeza.

2.2.4 Ambigüedad en la TA

Según explica Vázquez Ayora (1977), por un lado la ambigüedad se produce cuando falla el manejo del léxico y de la sintaxis; por otro lado, puede haber oraciones gramaticales que dan dificultades de interpretación. Eso no siempre vale para las ambigüedades en las traducciones automáticas. Las ambigüedades en las traducciones automáticas no siempre coinciden con las ambigüedades en la lengua humana, donde hay muy pocas oraciones que realmente sean ambiguas. La mayoría de las ambigüedades estructurales se podrían resolver con la información contextual o con el conocimiento del mundo real.

Como anteriormente hemos mencionado, muchas veces el contexto revela cuál de las posibles interpretaciones es la más correcta. En el ejemplo 11 (“El hombre vio a la niña con el telescopio”), probablemente se puede deducir la interpretación adecuada del contexto. Un traductor automático no es capaz de resolver este tipo de ambigüedades mediante la información contextual, lo que hace que el traductor automático se vea enfrentado con muchas más ambigüedades que el traductor humano. Además el traductor humano tiene conocimiento del mundo real, algo que el traductor automático no tiene. En las oraciones 17 y 18 encontramos el mismo tipo de ambigüedad estructural que en el ejemplo 11. Sin embargo, para un traductor humano estas oraciones no son ambiguas porque sabe que es improbable que *con el telescopio* se refiera a *caballo*, y por lo tanto modifica obligatoriamente a *vio*. Y en la frase 18 sabe que *con el pelo rojo* no puede referirse a *vio*, porque *pelo* no puede servir como un instrumento para ver. En cambio, para un traductor automático, que no tiene conocimiento del “mundo real”, estas dos oraciones sí son ambiguas.

17. El hombre vio el caballo con el telescopio.

18. El hombre vio a la chica con el pelo rojo.

En la práctica resulta imposible codificar e incorporar todo el conocimiento del mundo real necesario para resolver todas las posibles ambigüedades de un sistema concreto. En el capítulo 4 nos dedicaremos a este tipo de ambigüedades en los textos seleccionados para el experimento.

2.3 Conclusión

La complejidad de la TA está vinculada con la ingeniosidad de las lenguas naturales. Una lengua natural no sólo es muy precisa y especializada, sino también arbitraria y ofrece una infinita cantidad de construcciones. Cuando vinculamos la complejidad de las lenguas naturales con las limitaciones de los ordenadores, nos enfrentamos con muchos problemas a la hora de automatizar el proceso de traducción. Uno de los problemas más complejos es la ambigüedad. Incluso para un traductor humano la ambigüedad puede producir problemas de interpretación. No obstante, el traductor automático se enfrenta con muchas más ambigüedades que el traductor humano. Contrariamente a los traductores humanos, el traductor automático no puede deducir la interpretación adecuada del contexto y tampoco de su conocimiento del mundo real. En el tercer capítulo explicaremos la metodología que hemos usado en este estudio.

Capítulo 3

Metodología

En este capítulo expondremos la metodología que usamos para investigar el alcance de la traducción automática. Curiosamente, a pesar de los 60 años de investigación en la TA, todavía no se ha establecido una metodología generalmente aceptada para evaluar los sistemas de TA. En este capítulo discutiremos la complejidad de la evaluación de estos sistemas y explicaremos el método de investigación que hemos utilizado. Como la calidad de una TA no es unívoca, la enfocamos desde diferentes perspectivas. A través de un análisis de tipologías textuales pretendemos comparar la calidad de las traducciones automáticas de los diferentes tipos de texto. Comentaremos también la división textual y explicaremos la selección de los cuatro textos que hemos usado en el experimento.

3.1 Objetivo del estudio

En el presente trabajo buscamos respuesta a la pregunta: ¿cuál es el alcance actual de la TA en cuanto al par de lenguas español-neerlandés? El objetivo de nuestra de esta tesina será:

Determinar el alcance actual de la traducción automática en cuanto al par de lenguas español-neerlandés.

Es decir, pretendemos investigar la utilidad de la TA a la hora de traducir un texto. Por ello, no investigamos la utilidad de la TA como recurso para poder entender un texto (TA para la

comunicación), sino la utilidad de la TA como recurso para los traductores. Investigaremos el alcance sobre la base de una evaluación de cuatro textos españoles y sus respectivas traducciones al neerlandés, realizadas por un sistema de traducción automática. Nos centraremos en cómo el traductor automático traduce los diferentes tipos de textos y cómo procesa las complejidades inherentes a la traducción automática de una lengua natural. En particular, vamos a detenernos en cómo procesa las dificultades acerca de la ambigüedad.

3.2 La evaluación de la TA

La evaluación de la calidad del producto de la TA es el tema central en este estudio. Para poder determinar el alcance de la TA consideramos esencial comprobar la calidad del “producto crudo” de la TA, o sea, la calidad de la TA sin interferencias de un traductor humano. La evaluación de la TA siempre ha sido un tema central para los investigadores en este campo, pero todavía no existe una metodología generalmente aceptada para la evaluación de la TA. 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 consideración los mecanismos del traductor automático, sino que sólo analizamos el texto fuente y su respectiva traducción automática. Al contrario, con la visión de la “caja de vidrio” analizamos los mecanismos del traductor automático y evaluamos si todos los componentes técnicos funcionan correctamente.

En el presente estudio no tomamos en consideración las valoraciones técnicas y tampoco evaluamos los costos y beneficios de la TA, sino que optamos por el método de la “caja negra”, es decir, pretendemos investigar la calidad del producto crudo de un traductor automático.

La evaluación de la calidad de una traducción es una cuestión difícil por varios motivos. Para poder evaluar un texto traducido automáticamente necesitamos una norma, o sea, una traducción estándar, en otras palabras, necesitamos una traducción “correcta” o “ideal” con la cual podamos comparar nuestra traducción automática. Sin embargo, encontrar esta traducción ideal es imposible. Ningún documento es traducido de la misma manera por dos personas. Si evaluamos un documento traducido por varias personas, encontraremos traducciones de diferentes calidades sin que encontremos una traducción “ideal” y, además, los traductores probablemente no estarían de acuerdo sobre cuál de las traducciones sería la mejor. La evaluación de una traducción (y de una TA en particular) es en gran medida subjetiva.

En la bibliografía al respecto encontramos diferentes métodos para determinar la calidad de una TA. Una evaluación que tuvo consecuencias drásticas, es la publicada en el informe *ALPAC* en 1966 (véase el apartado 1.3). Los resultados negativos de esta evaluación provocaron que por muchos años se denegaran todos los fondos oficiales para investigaciones en la TA en los Estados Unidos. La evaluación de este informe consistía en medir la fidelidad, preguntando a los encuestados que leían una TA cuánto más “informativo” era el texto original. Esta manera de evaluar la podemos considerar como excesivamente subjetiva.

Una medida fácil para evaluar una TA es mediante la “retraducción”, lo que quiere decir que traducimos un texto con un traductor automático y después retraducimos el mismo texto a la lengua del texto fuente. De esta manera tenemos dos frases en la misma lengua y podemos observar fácilmente las diferencias entre ambas frases. Por ejemplo, cuando retraducimos con el traductor *Google* la frase: “Érase una vez, un borracho que abordó un autobús en el que viajaba mucha gente.”, la retraducción es: “Érase una vez, un borracho a bordo de un autobús que viajaba a mucha gente”. Queda claro que esta manera de evaluar tiene muchas deficiencias. En el proceso de la retraducción los errores se multiplican, y por tanto, no dice mucho sobre la calidad. Una medida frecuentemente usada en la evaluación de la TA es el *Words Error Rate (WER)*, es decir, el número de errores dividido por el total de palabras del texto. Cuanto menor sea el índice, mejor es la traducción. Obviamente, esta manera de evaluación también tiene sus puntos débiles. Muchas veces los errores se encuentran en diferentes niveles, con la consecuencia de que no se puede detectar un error por palabra. Además, es difícil calificar algo como un error de traducción, dado que en una traducción no siempre se puede hablar objetivamente de mal o bien.

3.3 El método seleccionado

Como anteriormente hemos mencionado, en el presente estudio utilizamos un método de “caja negra” para evaluar la calidad de la TA. Aplicamos un método que en la literatura se denomina “evaluación declarativa”. Hemos optado por este tipo de evaluación porque estamos particularmente interesados en la capacidad del sistema actual y no en su extensión potencial. Como ya hemos mencionado, es difícil determinar cuándo es mala, buena o excelente la calidad de una traducción. Puesto que es imposible ignorar la subjetividad en la evaluación, intentamos aprovechar de esta misma subjetividad en forma de la intuición lingüística. En el presente estudio distinguiremos entre una evaluación cuantitativa y otra cualitativa.

3.3.1 Evaluación cuantitativa

En este estudio evaluamos la calidad de la TA a través de las opiniones de expertos, que denominamos “evaluación cuantitativa”. Los expertos evalúan en una escala de cinco puntos la calidad de una oración traducida automáticamente con la presencia de su oración original (véase la figura 3.1.). La parte izquierda significa que se trata de una mala traducción y la parte derecha significa que se trata de una buena traducción. Este método, establecido por Krings (2001) para comparar diferentes sistemas de TA, nos permite explotar la subjetividad que es inherente a la evaluación de la TA en forma de la intuición lingüística. Aunque los encuestados no siempre estén de acuerdo sobre qué es exactamente una traducción correcta, en este cuestionario sólo deben seguir sus intuiciones lingüísticas. Cada persona puede dar fácilmente un juicio de valor sobre cuándo suena mal o bien una frase. Por este motivo no hemos definido lo que es “goed” o “slecht”, sino que

confiamos en las intuiciones lingüísticas del evaluador. Suponemos que los elementos de fidelidad, exactitud, fluidez y comprensibilidad juegan un papel importante, pero no investigamos estos elementos explícitamente. Con este cuestionario el encuestador puede evaluar una gran cantidad de frases traducidas con relativamente poco esfuerzo. La media aritmética de cada oración valorada representa su calidad y nos posibilita dar una valoración de la TA. Al final del cuestionario los expertos pueden dar su impresión general de las traducciones en una pregunta abierta.

F1: Asistir a una actividad
 F1: Voor het bijwonen van een evenement
 slecht goed

F2: El centro de Utrecht organiza debates, conferencias, presentaciones de libros, encuentros de escritores, artistas, profesionales o científicos,
 F2: *Het centrum van Utrecht georganiseerde debatten, lezingen, boekpresentaties, ontmoetingen met schrijvers, kunstenaars, professionals en wetenschappers,*
 slecht goed

F3: siempre con traducción simultánea español-neerlandés y a la inversa.
 F3: *mits met gelijktijdige Spaans-Nederlands en vice versa.*
 slecht goed

F4: También ofrece música, teatro, cine y exposiciones de formato pequeño,
 F4: *Ook biedt muziek, theater, film en tentoonstelling van klein formaat,*
 slecht goed

Figura 3.1 Formulario de evaluación de la calidad de la TA

Encuestados

Las personas encuestadas en este experimento son los usuarios potenciales de la TA. Es esencial para este experimento que los encuestados conozcan la lengua española y la neerlandesa, y que entiendan el proceso de traducción. El grupo encuestado consiste en veinte personas, de las que quince son estudiantes del Máster de Traducción y cinco son traductores profesionales. La opinión de este grupo es particularmente interesante porque serán los usuarios potenciales. Consideramos a este grupo de encuestados como “expertos” y por lo tanto sus valoraciones de la calidad de la TA nos parecen representativas.

3.3.2 Evaluación cualitativa

Con la evaluación cualitativa nos referimos a la evaluación lingüística del producto crudo de la TA, o sea, a la clasificación de los errores lingüísticos. Analizamos los errores lingüísticos en los textos seleccionados. No usamos un modelo estándar de clasificación de los errores, sino que clasificamos los errores a partir de los textos. En el caso de que aparezca más de un error, cada error vale. Por ejemplo, un sustantivo puede tener una traducción semántica errónea y, además, puede tener una

posición sintáctica errónea, en cuyo caso contaríamos dos errores. Estamos particularmente interesados en los errores producidos por la ambigüedad. Aunque hacemos esta evaluación independientemente de la evaluación cuantitativa, sí analizamos las relaciones entre ambos estudios. La información sobre la calidad media de cada frase y la categoría lingüística a la que pertenece, nos permite contestar a las siguientes preguntas relevantes: ¿qué tipo de error encontramos en qué tipo de texto? ¿Qué tipo de error contienen las oraciones peor evaluadas? ¿Qué tipo de error contienen las oraciones mejor evaluadas?

Post-edición

La post-edición, es decir, la revisión y corrección del texto traducido es un proceso que también se realiza en las traducciones humanas. Opinamos que el grado de post-edición es inherente a la calidad de la TA. En otras palabras, lo que hay que revisar más, es de peor calidad. Obviamente existen diferentes niveles de post-edición. Algunos errores se pueden corregir fácilmente (como por ejemplo el orden erróneo de palabras), mientras que otros errores son más difíciles de corregir (errores pragmáticos). Llamamos “errores fáciles de revisar” a los errores que se pueden corregir sin texto fuente y, en cambio, llamamos “errores difíciles de revisar” a los errores que no se pueden corregir sin texto fuente. Para extraer los “errores difíciles de revisar”, dos traductores profesionales revisaron una TA sin texto fuente.

Analizaremos estas traducciones revisadas y evaluaremos los errores que hemos encontrado en el texto, o sea, los errores difíciles de revisar.

Traductor automático

En este experimento nos concentramos en la aplicación traductora de *Google*, a saber, *Traductor Google*. Hemos optado por este traductor automático porque no usa un método basado en normas como la mayoría de los sistemas comerciales, lo cual requiere que todo el vocabulario y las reglas gramaticales de una lengua estén definidos, sino que usa un método estadístico (véase el apartado 1.2.2). A pesar de que el método estadístico no ha resultado mejor que el método basado en normas, opinamos que este método es el más innovador y prometedor. Además, el *Traductor Google* es uno de los pocos traductores automáticos que ofrecen el par de lenguas español-neerlandés.

3.3.3 Debilidades

Un experimento nunca es infalible cien por cien, y para obtener el máximo valor de un estudio es esencial conocer sus debilidades. Como en este estudio usamos el método de la “caja negra”, no podemos sacar conclusiones precisas sobre la capacidad del sistema. El traductor automático puede haber producido “accidentalmente” las traducciones exactas de los textos seleccionados. Cuatro textos no son suficientes para sacar conclusiones precisas sobre la capacidad del traductor automático. Además, podemos encontrar algunos problemas a la hora de rellenar el cuestionario. Aunque ofrecemos en el cuestionario el texto fuente y su respectiva TA completa, el encuestador da un juicio de valor sobre las oraciones aisladas. Una oración aislada puede causar una valoración más

negativa que en su contexto. También tenemos que tener en cuenta que los expertos no evalúan de la misma manera. Los resultados indican que los traductores profesionales generalmente evalúan más negativamente que los estudiantes del Máster de Traducción. Además, el orden de las oraciones puede influir, después de evaluar una oración muy mala, la siguiente oración puede parecer mejor. Estas debilidades son inherentes a la evaluación de la TA. A pesar de eso, tenemos suficiente confianza en el valor de este estudio, que, aunque sea limitado en ciertos puntos, creemos que proporciona unos resultados suficientemente interesantes para iniciar más investigaciones.

3.4 La selección de los textos

Hacemos el experimento sobre cuatro textos que hemos seleccionado basándonos en la tipología textual de Reiss (1976), que se basa en la función comunicativa del texto. El autor identifica tres tipos de texto, a saber, la función informativa, expresiva y operativa. Reiss opina que la equivalencia está en la función comunicativa del texto y, por lo tanto, vincula cada tipo de texto a una estrategia de traducción específica.

Texto informativo: el foco principal es comunicar el contenido como ocurre en los textos científicos y técnicos. El tema determina las decisiones a seguir respecto a su verbalización y, por lo tanto, su traducción debería transmitir la información contenida en el texto fuente en su totalidad sin redundancia y con explicaciones cuando sean necesarias. La dimensión del lenguaje para transmitir la información es lógica y referente.

Texto expresivo: su rasgo característico es el valor emotivo y artístico. El elemento principal en la situación comunicativa es el emisor que explota las posibilidades de la lengua para crear el texto. El emisor, o autor, está en primer plano, así como la forma del mensaje. Utiliza la dimensión estética de la lengua. El principal objetivo del traductor ha de ser, por tanto, crear otro texto en la lengua meta con la misma función artística, adaptándose al punto de vista del autor del texto fuente.

Texto operativo: su objetivo es atraer la atención del receptor o inducirle a realizar una acción determinada como acontece en los anuncios publicitarios o en los discursos políticos. La traducción de este tipo de texto debería producir sobre el receptor del texto meta el mismo efecto que tuvo el texto fuente en su receptor⁵.

Como podemos observar, el concepto de equivalencia que plantea Reiss no propone una equivalencia en palabras o frases, sino en la función comunicativa del texto. Propone una estrategia de traducción específica para cada tipo textual. En la tabla 3.1 resumimos las características de cada tipo textual y su respectiva estrategia de traducción.

⁵ Reiss propone un cuarto tipo de texto, el audiomedial. Este tipo incluye los textos acompañados de otros signos no-verbales (como música, imágenes, etcétera.). Como en este estudio sólo abordamos textos escritos no tomamos en consideración este tipo de texto.

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

Tabla 3.1 Tipología textual (Reiss 1976)

Reiss (1976) asigna una única función a cada texto. Esta rigidez no la encontramos cuando colocamos en la pirámide de Chesterman (1997) los textos seleccionados. Ahora podemos observar que muchas veces los textos no son únicamente informativos, expresivos u operativos, sino que tienen una combinación de estas funciones. Por ejemplo, un discurso político no es meramente operativo, también puede funcionar como informativo y expresivo. Es decir, no sólo quiere incitarse al electorado, también se quiere informar al público de una forma estética. El “yes we can” del presidente Obama es más que sólo operativo, también tiene algo estético. En la figura 3.2 presentamos los textos que hemos seleccionado para el experimento. Para poder formular una hipótesis sobre qué función comunicativa es la más difícil de traducir automáticamente, hemos seleccionado cuatro textos de tres tipologías textuales diferentes.



Figura 3.2 Pirámide de Chesterman con las funciones del texto

Primer texto: Instituto Cervantes

El primer texto⁶ procede del sitio web español del *Instituto Cervantes*, una institución pública creada por España para la promoción y la enseñanza de la lengua española y para la difusión de la cultura española e hispanoamericana. Este texto se encuentra en la página que contiene información general sobre el Instituto. Trata de las actividades culturales que se ofrecen y da información sobre las actividades culturales y cómo se puede participar en ellas. Está escrito de forma neutral y atractiva. En primer lugar quiere informar a la gente y en segundo lugar quiere atraer a la gente. Por lo tanto, lo hemos clasificado como principalmente “informativo”.

Segundo texto: folleto turístico

El segundo texto⁷ es un texto turístico sobre Barcelona publicado en el sitio web oficial sobre turismo de Barcelona. El objetivo del texto es obvio: atraer a los turistas para que vayan a Barcelona para pasar la primavera. Además de atraer al lector, el escritor quiere informar sobre las posibilidades de Barcelona. El texto está escrito en un lenguaje atractivo para seducir al lector. Lo consideramos principalmente como ‘operativo’. A la hora de traducir, hay que mantener este lenguaje atractivo para que el texto produzca el mismo efecto en la cultura meta.

Tercer texto: Receta

El tercer texto⁸ es una receta de tortilla de patatas. En este texto encontramos instrucciones para preparar tortilla. El texto consiste en instrucciones y sugerencias, escritas en una lengua neutral. Es un texto evidentemente informativo. A la hora de traducir este texto, hay que transmitir los hechos, para que el lector del texto meta entienda qué tiene que hacer para preparar la misma tortilla de patatas que el escritor del texto fuente.

Cuarto texto: chiste

El cuarto texto⁹ es un chiste que proviene de un sitio web que publica todo tipo de chistes. Trata de un borracho en un autobús que llama a los pasajeros idiotas, imbéciles y estúpidos. La broma es que después de que el conductor enfadado había frenado y toda los pasajeros habían caído al suelo, el borracho dijo que ya no podía distinguir a los tarados, idiotas e imbéciles porque estaban *todos revueltos*. Consideramos este texto principalmente como expresivo. A la hora de traducir es muy importante transmitir la forma del texto según “las normas” de un chiste en la lengua meta. Por ello, para transmitir el mismo efecto, no es suficiente traducir el contenido solamente.

⁶ El texto completo se encuentra en el anexo 1.

⁷ El texto completo se encuentra en el anexo 2.

⁸ El texto completo se encuentra en el anexo 3.

⁹ El texto completo se encuentra en el anexo 4.

3.5 Conclusión

A pesar de los 60 años de investigación, todavía no existe una metodología generalmente aceptada para la evaluación de la TA. Como estamos particularmente interesados en el producto crudo del traductor automático utilizamos el método de la “caja negra”. Mediante una evaluación cualitativa y cuantitativa determinaremos el alcance actual de la TA del par de lenguas español-neerlandés. Evaluaremos el producto crudo y clasificaremos los errores lingüísticos. Prestaremos atención especial a los errores producidos por la ambigüedad. Haremos énfasis en la utilidad de la TA como recurso del traductor, y por ello, más que nada, partiremos de las opiniones de los usuarios potenciales sobre la calidad de la TA. Como es muy complejo determinar objetivamente cuándo es mala o buena la calidad de una TA, usaremos la intuición lingüística del usuario potencial. Sus valoraciones nos permitirán clasificar las oraciones y sacar conclusiones sobre la calidad de los diferentes textos traducidos automáticamente. Hemos seleccionado cuatro textos basándonos en la división de Reiss (1976), quien propone una equivalencia en la función comunicativa del texto, la cual requiere una estrategia de traducción específica para cada tipo textual. Puede ser que al igual que ocurre con un traductor humano, el traductor automático sea mejor para traducir un texto informativo que para un texto operativo o expresivo. En el capítulo 4 presentaremos los resultados de nuestros experimentos.

Capítulo 4

¿Cuál es el alcance?

En este capítulo presentaremos los resultados del experimento e intentaremos contestar a la pregunta: ¿cuál es el alcance actual de la traducción automática respecto al par de lenguas español-neerlandés? Para determinar el alcance, tenemos que conocer la calidad del “producto crudo” de la TA. Como la calidad de la TA no es unívoca, abordaremos la calidad de la TA desde diferentes perspectivas. Primero evaluaremos los textos desde el punto de vista lingüístico, es decir, determinaremos la calidad subrayando los errores lingüísticos en cada frase. Prestaremos atención especial a la aparición de ambigüedades en los textos. A continuación presentaremos las valoraciones de los expertos, es decir, presentaremos las oraciones y los textos mejor y peor evaluados. Con la ayuda de la evaluación lingüística intentaremos descubrir por qué los expertos evalúan una oración mejor que otra. En otras palabras, relacionaremos los errores lingüísticos con las valoraciones de los expertos. Por último, analizaremos la calidad después de la post-edición del primer texto.

4.1 La evaluación lingüística

En este apartado evaluaremos los errores lingüísticos en los cuatro textos. Como ya hemos explicado, no usamos un modelo estándar para clasificar los errores, sino que clasificaremos los errores a partir de los cuatro textos. Por ejemplo, la aparición del sujeto en español difiere mucho del

neerlandés, lo cual ocasiona errores de este tipo. Por ello, distinguimos las categorías “sujeto erróneo” y “omisión del sujeto”. Además, hemos subdivido la categoría “omisión”, porque encontramos muchas omisiones de, entre otras cosas, verbos y sujetos. Como hemos seleccionado cuatro textos diferentes, investigaremos si hay una relación entre la clase de errores y el tipo de texto. En cada texto evaluaremos los errores y los pondremos en orden de más a menos frecuentes. Distinguiremos las siguientes categorías de errores:

- Léxico: palabra errónea
- Léxico: ambiguo
- Léxico: otros
- Función gramatical errónea
- Sintáctico: sujeto erróneo
- Sintáctico: orden de palabras
- Sintáctico: otros
- Morfológico
- Omisión del verbo
- Omisión del sujeto
- Omisión: otros
- Preposición errónea
- Varios

4.1.1 Evaluación lingüística del texto 1

El primer texto trata sobre las actividades del *Instituto Cervantes*; lo consideramos principalmente informativo. El texto original se encuentra en el anexo 1. En este texto hay un error cada 5,7 palabras.

[F1] Voor het bijwonen van een evenement

[F2] Het centrum van Utrecht georganiseerde debatten, lezingen, boekpresentaties, ontmoetingen met schrijvers, kunstenaars, professionals en wetenschappers, [F3] mits met gelijktijdige Spaans-Nederlands en vice versa. [F4] Ook biedt muziek, theater, film en tentoonstelling van klein formaat, [F5] zowel in het centrum en in de andere ruimten en vaak in samenwerking met andere culturele entiteiten.

[F6] Culturele activiteiten meestal plaats in het centrum in de middag. [F7] Gewoonlijk is de entree is 5 euro, [F8] met een korting van \$ 2 voor studenten ouder dan 65 en de leden van de bibliotheek. [F9] Onze studenten betalen geen toegang. Veel van de activiteiten zijn gratis.

[F10] Meestal is het publiek heeft de kans om vragen te stellen na de gebeurtenis [F11] en een praatje met de gasten onder het genot van een Spaanse wijn. [F12] Het is raadzaam om een plaats in alle gevallen reserveren door te bellen naar 030 24 28 473 [F13] of schrijven naar onze culturele afdeling: cultutr@cervantes.es

[F14] Wie geïnteresseerd is kan ontvangen ieder kwartaal een papier en / of een elektronische nieuwsbrief met een herinnering van elke activiteit.

Texto 1 *Actividades en el Instituto Cervantes*

1. Léxico: ambiguo

La mayoría de los errores de este texto pertenece a la categoría “ambigüedad léxica”. Como hemos explicado en el apartado 2.2.1 la palabra “centro” (la frase 1) es una palabra polisémica, es decir, muestra una variedad de significados relacionados de algún modo entre sí. Esta polisemia existe tanto en la lengua fuente como en la lengua meta, pero opinamos que en este contexto, para evitar la ambigüedad, sería mejor traducirlo por “instituut” o “*Instituto Cervantes*”. La palabra “entidad” (la frase 5) produce un mismo tipo de ambigüedad, tanto si se refiere al significado filosófico “entiteit”, como a “instelling”. En neerlandés es más común usar en este contexto la palabra “instelling”.

2. Léxico: otros

Otra categoría de errores muy frecuente, es la categoría con los “otros” errores léxicos. A esta categoría pertenecen, entre otras, las palabras “añadidas”. Por ejemplo en la primera frase la preposición “voor” en la traducción de “Asistir a una actividad” no es necesaria. El error que encontramos en la frase 5 también lo colocamos en esta categoría. En esta frase se traduce “tanto [...] como” como “zowel [...] en”, en lugar de “zowel [...] als”. Es una colocación fija en neerlandés.

3. Sintáctico: orden de palabras

El orden de palabras erróneo también aparece mucho en este texto. En la frase 6 la posición de “in het centrum” no es correcta en la frase “Culturele activiteiten meestal plaats in het centrum in de middag” e incluso tiene consecuencias para la semántica de la frase. En la frase original “Las actividades culturales en el centro suelen tener lugar por la tarde”, la parte “en el centro” pertenece al sujeto.

4. Función gramatical

A este tipo de errores pertenecen las palabras que tiene una función gramatical errónea. Por ejemplo un sustantivo que tiene que ser un adjetivo. En la frase 2 encontramos un adjetivo que tiene que ser un verbo. El verbo “organiza” en la frase original “El centro de Utrecht organiza debates [...]” es traducido por un adjetivo en el texto meta: “Het centrum van Utrecht georganiseerde debatten [...]”.

5. Varios

Esta categoría contiene los errores que no pertenecen a las otras categorías. En este texto encontramos tres de estos errores. Por ejemplo, en la frase 8 se traduce el símbolo del euro por el del dólar. Y en las frases 12 y 14 falta la palabra “te” antes del infinitivo, lo que influye en la fluidez de la frase.

6. Léxico: palabra errónea

A esta categoría pertenecen los errores respecto a la traducción del léxico. Por ejemplo, en la frase 1 se traduce “actividad” por “evenement”, mientras que la traducción correcta sería “activiteit”. Y en la frase 3 “siempre” ha sido traducido por “mits” en lugar de “altijd”.

7. Omisión: verbo

En esta categoría encontramos los verbos no traducidos. Por ejemplo en la frase 6 la parte “[...] suelen tener lugar [...]” ha sido traducida por “[...] meestal plaats [...]”. Lo que falta en esta frase es el verbo “vinden”, para la traducción correcta de “tener lugar”.

8. Omisión: otros

A esta categoría pertenecen las omisiones que no son sujetos o verbos. En la traducción de la frase 3 “siempre con traducción simultánea español-neerlandés y a la inversa”, el sustantivo “traducción” no se encuentra en el texto meta: “mits met gelijktijdige Spaans-Nederlands en vice versa.”

9. Sintáctico: otros

A esta categoría pertenecen los “otros” errores sintácticos. Por ejemplo, en la frase 8 encontramos un error con respecto a la puntuación. La frase original “con un descuento de 2 euros para estudiantes, mayores de 65 [...]” se traduce como “met een korting van \$ 2 voor studenten ouder dan 65 [...]”. En este caso, un error en la puntuación tiene consecuencias para la semántica, porque ahora parece que sólo los estudiantes mayores de 65 años reciben el descuento, mientras que el texto fuente se dirige a los estudiantes y a los mayores de 65.

10. Omisión: sujeto

El sujeto en la lengua fuente tiene otra forma de aparición que en la lengua meta. El español muchas veces tiene un sujeto implícito, mientras que el neerlandés suele tener un sujeto explícito. La frase 2 no funciona sin el sujeto: “Ook biedt muziek, theater, film en tentoonstelling van klein formaat”.

11. Preposición errónea

La traducción de preposiciones es una cuestión difícil. En la frase 10 el traductor automático traduce “al término del acto” por “na de gebeurtenis”, lo que no es exactamente lo mismo. Mejor sería traducirlo por “aan het einde van de activiteit”.

4.1.2 Evaluación lingüística texto 2

El segundo texto es un texto turístico sobre la primavera en Barcelona. Lo consideramos como principalmente operativo. El texto original se encuentra en el anexo 2. En este texto aparece un error cada 4,3 palabras.

[F1] Wat is hier? En sommige straten verderop? [F2] Waar u ook bent in Barcelona, heel dichtbij, de buurt of wijk, is er iets te zien: [F3] de Art Nouveau sieraden en hedendaagse architectuur, [F4] de markten voor de zintuigen genieten van de schatten van de oude Romeinse en middeleeuwse stad, [F5] parken ontspannen ... [F6] En het beste deel is dat er een grote ontdekkingsreiziger te vinden en kennen de buurten van Barcelona.

[F7] Dit voorjaar in Barcelona

[F8] Barcelona biedt interessante mogelijkheden voor alle seizoenen. [F9] Hier zijn enkele voorstellen gevarieerd:

[F10] "Een gastronomische wandeling
"Alle kunst van Barcelona
Bezoeken in comfort-

[F11] "Een magisch-kaart
Gitaar-en architectuur

[F12] Maar er is zoveel meer. [F13] Neem de zon schijnt voor wandelingen in de parken en stranden. [F14] Ontdek de belangrijkste tentoonstellingen en neemt te doen uw boodschappen. [F15] En ontdek de keuken van het seizoen, [F16] en genieten van een concert, [F17] en de ervaring van de aard van Gaudí? [F18] Wil je meer ideeën?

Texto 2 Folleto turístico

1. Preposición errónea

La mayoría de los errores en este texto está relacionada con la traducción de las preposiciones. Muchas veces el traductor automático no reconoce la preposición como tal y la omite o la traduce por una preposición neerlandesa errónea. Por ejemplo, en la frase 3 "joyas del Modernismo y de la arquitectura contemporánea", no se ha traducido la preposición "de" correctamente. Se ha convertido "joyas del Modernismo" en una palabra compuesta "Art nouveau sieraden". Obviamente, aquí no se trata de una palabra compuesta, porque "joyas" se refiere tanto al Modernismo como a la arquitectura. Y en la frase 4 se ha traducido "para" en la frase "mercados para regalarse los sentidos" por "voor". En este contexto "para" no tiene el sentido de "voor" sino de "om".

2. Omisión: otros

En este texto muchas palabras están omitidas, lo cual puede tener consecuencias para la semántica. Por ejemplo, en la frase 6 se ha omitido la palabra "no" en "Y lo mejor es que no hay que ser un gran explorador [...]". Obviamente el significado de la frase cambia, de una frase negativa a una positiva. En la frase 13 se omite la parte "de Barcelona" en "Aprovecha que el sol brilla para pasear por los parques y playas de Barcelona".

3. Léxico: palabra errónea

En este texto encontramos muchos errores léxicos. Por ejemplo, la traducción de "aprovechar" por "nemen" en la frase 13 "Aprovecha que el sol brilla [...]". En esta frase "aprovechar" tiene el significado de "genieten".

4. Léxico: palabra ambigua

A esta categoría pertenecen las palabras ambiguas, como, por ejemplo, la expresión "hacer compras" en la frase 14. "Hacer compras" se refiere tanto a "boodschappen doen" como a "winkelen", que en neerlandés son dos conceptos distintos. "Boodschappen doen" se refiere a hacer compras en un supermercado y "winkelen" se refiere a hacer compras en, por ejemplo, tiendas de moda. Lo más lógico sería que en este texto se refiera al concepto "winkelen".

5. Sintáctico: sujeto erróneo

A esta categoría pertenecen los verbos que se refieren a un sujeto erróneo. En este texto encontramos algunos ejemplos de este fenómeno. En la frase 5 se traduce "parques donde relajarse..." como "parken ontspannen". En la traducción automática "parken" es el sujeto, mientras que en el texto fuente se trata de una frase impersonal que automáticamente no tiene sujeto explícito.

6. Morfológico

Los errores morfológicos que aparecen en este texto se encuentran, por ejemplo, en la frase 14 "Ontdek de belangrijkste tentoonstellingen en neemt te doen uw boodschappen." El verbo "ontdekken" está conjugado correctamente como imperativo, sin embargo "nemen", que debería estar conjugado igualmente, está conjugado como un presente del indicativo. Otro error morfológico se encuentra en la oración 11 "Een magisch-kaart [...]". La forma del adjetivo "magisch" tiene que ser "magische".

7. Función gramatical

En este texto encontramos algunos errores respecto a la función gramatical. Por ejemplo en la frase 17 se traduce el verbo "vive" en la frase "y vive la naturaleza de Gaudí" por "ervaring" ("en de ervaring van de aard van Gaudí"). La traducción correcta sería usar el verbo "ervaren".

8. Sintáctico: orden de palabras

Encontramos por ejemplo algunos errores sintácticos con respecto al orden en la frase 9: "Hier zijn enkele voorstellen gevarieerd:". El adjetivo "gevarieerd" debería estar antes del sustantivo "voorstellen". El orden de palabras erróneo tiene consecuencias para la fluidez de la frase.

9. Sintáctico: otros

En esta categoría encontramos los "otros" errores sintácticos, es decir los errores sintácticos que no están relacionados con el orden de palabras. La frase 13 consiste en una oración subordinada "Aprovecha que el sol brilla [...]". La traducción automática no usa la misma construcción, lo que ocasiona una oración incomprensible. Otro error sintáctico es el que encontramos en la frase 4. El traductor automático no ha reconocido la coma, lo que ocasiona una traducción errónea de "mercados para regalarse los sentidos, tesoros de la antigua ciudad romana y medieval" por "de markten voor de zintuigen genieten van de schatten van de oude Romeinse en middeleeuwse stad,". Como no hay una coma, en la traducción el sujeto es, "de markten voor de zintuigen".

10. Omisión: sujeto

En la TA de la frase 6 no hay un sujeto “En het beste deel is dat er een grote ontdekkingsreiziger te vinden en kennen de buurten van Barcelona”. En el texto fuente hay un sujeto impersonal “Y lo mejor es que no **hay que ser** un gran explorador para encontrarlos y conocer todos los barrios de Barcelona”. En la parte “hay que ser” el sujeto es impersonal, lo que se puede traducir en neerlandés como “je” o “men”.

11. Omisión: verbo

En la misma frase 6, también se ha omitido el verbo “ser”, “Y lo mejor es que no hay que **ser** un gran explorador [...]”. El resultado de la frase es incomprensible: “En het beste deel is dat er een grote ontdekkingsreiziger te vinden en kennen de buurten van Barcelona.”

12. Varios

Esta categoría contiene los errores que no pertenecen a las otras categorías. La frase 18 “¿necesitas más ideas?” es traducida literalmente como “wil je meer ideeën?”. Aunque la frase traducida automáticamente es gramaticalmente correcta, no tiene el mismo efecto que la frase original. Pragmáticamente sería más adecuado traducirla como “wil je **nog** meer ideeën?” o “heb je nog meer ideeën nodig?”.

4.1.3 Evaluación lingüística texto 3

El tercer texto es una receta de tortilla de patatas. Consideramos este texto principalmente como informativo. El texto original se encuentra en el anexo 3. En este texto se encuentra un error cada 4,5 palabras.

[F1] Recept voor aardappel tortilla

[F2] Schil en was de aardappelen [F3](naar schatting een grote aardappel per persoon). [F4] Het snijd ze in plakjes niet te groot.

[F5] Bak de aardappelen in ruim olie. [F6] Om de tortilla is erg belangrijk de aardappelen zijn zacht en niet knapperig. [F7] Dit wordt gedaan door frituren warmte laag [F8] en in een afgedekte pan voor ongeveer 10 minuten.

[F9] In een grote kom (bijvoorbeeld een schaal) Klop de eieren [F10] (het Naar schatting een ei per persoon). [F11] Voeg wat zout.

[F12] Wanneer de aardappelen zijn al zacht, te verwijderen uit de pan en afvoer van olie. [F13] Dan is goed te mengen met de losgeklopte eieren en laat het staan ongeveer 5 minuten.

[F14] In een ondiepe pan worden gegoten 2 eetlepels olie. [F15] Wanneer warm, giet het mengsel van aardappelen, eieren en ui en gebakken boven vuur medium.

[F16] Na een paar minuten in de koekenpan, draai je de tortilla. [F17] Wanneer de tortilla wordt gemaakt door beide partijen, kan worden gegeten. [F18] Bon Appetit!

Texto 3: Receta tortilla de patatas

1. Sintáctico: orden de palabras

La mayoría de los errores en este texto son errores con respecto a un orden de palabras erróneo. Por ejemplo en la frase 4 "Het snijd ze in plakjes niet te groot." En neerlandés el adjetivo normalmente precede al sustantivo, la traducción correcta de esta frase sería "in niet te grote plakjes".

2. Morfológico

En este texto encontramos muchos errores respecto a la conjugación de los verbos. Por lo general, no afecta tanto la comprensibilidad del texto, pero sí afecta la fluidez. En las primeras frases se usa un imperativo traducido en el texto, pero a partir de la frase 12 empieza a usarse un infinitivo. Por ejemplo, en la frase 13 se traduce "Entonces se mezclan bien [...]" por "Dan is goed te mengen [...]". Lo mejor sería usar el imperativo.

En la frase 5 encontramos un error con respecto a la forma del adjetivo "ruim". "Bak de aardappelen in ruim olie." La forma correcta sería "ruime".

3. Léxico: palabra ambigua

En este texto encontramos muchas palabras ambiguas, o sea, palabras con más de un significado. El traductor automático muchas veces optó por el significado erróneo. Por ejemplo en la frase 7, se traduce "freír" por "frituren", mientras que aquí se refiere a "bakken". Y en la frase 12 se traduce "las patatas ya blandas" como "de aardappelen zijn al zacht". En este contexto "ya" no tiene el significado de "al", sino de "net".

4. Varios

En el texto traducido encontramos palabras en otros idiomas, para las cuales ya existe una traducción en neerlandés. Por ejemplo en la frase 18 se traduce "¡Buen provecho!" como "Bonne Appetit!" y en la frase 15 se traduce "se fríe a fuego medio" como "[...] gebaken boven vuur medium". En ambos casos el neerlandés ofrece posibles traducciones correctas, a saber, respectivamente "eet smakelijk" y "half vuur".

También encontramos un error con respecto a un fenómeno muy común en la lengua neerlandesa, a saber, "los verbos separables". Estos son los verbos que se pueden separar, como por ejemplo el verbo "toevoegen" que consiste en la parte "voegen" y "toe". En la frase 11 "se echa sal", está traducido como "Voeg wat zout". Obviamente, aquí trata del verbo separable "toevoegen", y aquí falta la parte "toe".

5. Léxico: palabra errónea

En las frases 13 y 15 se traduce "sartén" como "pan" en lugar de "koekenpan". En neerlandés "pan" es otro utensilio de cocina, es una "olla". Sorprendentemente en la frase 16 sí se traduce "sartén" como "koekenpan". Otro error de este tipo es la traducción de "fuego" por "warmte" en la frase 7 "Para eso hay que freírlas a fuego bajo".

6. Omisión: sujeto

Algunas frases traducidas no tienen un sujeto, por ejemplo la frase 7 "Dit wordt gedaan door frituren warmte laag". La frase original es impersonal "Para eso hay que freírlas a fuego bajo". En neerlandés este tipo de frases requieren que se añada "men" o "je".

7. Omisión: otros

En el texto fuente hay algunas palabras que no se han traducido. Por ejemplo en la frase 7 no se traduce el pronombre "las" en la frase "Para eso hay que freírlas [...]".

8. Preposición errónea

En este texto se encuentran algunas preposiciones erróneas, por ejemplo en la frase 17 "Cuando la tortilla está hecha por los dos lados" se traduce "por" como "door" en lugar de "aan".

9. Léxico: otros

En la frase 4 encontramos una palabra 'añadida': "Het snijd ze in plakjes niet te groot." En esta frase "het" no tiene función. Lo mismo se puede apreciar en la frase 10 "(het Naar schatting een ei per persoon)." De nuevo aparece el artículo "het" que no desempeña ninguna función en la frase traducida.

10. Omisión: verbo

En la frase 6 "Para hacer la tortilla es importante que [...]" el verbo "hacer" no se ha traducido "Om de tortilla is erg belangrijk[...]".

11. Sintáctico: otros

La frase 6 consiste en una oración principal y una subordinada "Para hacer la tortilla es muy importante que las patatas estén blandas y no crujientes.". La frase traducida no tiene esta construcción, lo que resulta en una traducción que no es fluida: "Om de tortilla is erg belangrijk de aardappelen zijn zacht en niet knapperig."

4.1.4 Evaluación lingüística texto 4

El cuarto texto es un chiste. Consideramos este texto principalmente como expresivo. El texto original se encuentra en el anexo 4. Encontramos un error cada 2,9 palabras.

[F1] Once upon a time, [F2] een dronkelap die aan boord van een bus die op reis was een heleboel mensen, [F3] en staande in het gangpad van de bus, ging om te zeggen:
[F4] Die aan de rechterkant zijn een paar idioten, die aan de linkerzijde zijn idioten, [F5] de idioten zijn en achterkant zijn dom. [F6] Toen hij hoorde dat de bestuurder plotseling remmen, [F7] en iedereen viel op de grond, met inbegrip van dronken [F8] en heel erg boos de bestuurder nam de dronken bij de kraag en vroeg:
[F9] Nu, vertel me, wie zijn sommige idioten, idioten, imbecielen en dom?
[F10] En de dronken antwoordde:
[F11] Ik weet zelfs niet weet, worden alle gecodeerde!

Texto 4 Chiste

1. Léxico: palabra errónea

La mayoría de los errores en este texto pertenece a esta categoría. Por ejemplo, en la frase 11 “Ya ni sé, ¡están todos revueltos!” se traduce “revueltos” como “gecodeerde” y “estar” por “worden”, y resulta que esta frase es totalmente incomprensible.

2. Sintáctico: orden de palabras

También encontramos muchos errores con respecto al orden de palabras. Por ejemplo, la frase 6 “Cuando escuchó eso el conductor, frenó sorprendentemente,” se traduce como “Toen hij hoorde dat de bestuurder plotseling remmen”. En este caso, la posición errónea del sujeto “bestuurder” en la traducción tiene consecuencias para la semántica de la frase.

3. Léxicos: función gramatical

En este texto encontramos muchas palabras que desempeñan una función gramatical errónea. Por ejemplo, en la frase 7 se traduce “ el borracho” con el adjetivo “dronken” y en la frase 9 se traduce “unos estúpidos” con el adjetivo “dom”.

4. Morfológico

Este tipo de error se encuentra, por ejemplo, en la frase 6 “Cuando escuchó eso el conductor, frenó sorprendentemente”. La conjugación del verbo “remmen” no es correcta “Toen hij hoorde dat de bestuurder plotseling remmen”.

5. Léxico: palabra ambigua

Este error aparece, por ejemplo, en la traducción de la frase 11 “Ya ni sé [...]”. El traductor automático traduce “ni” como “zelfs” en “Ik weet zelfs niet [...]”. En esta oración la conjunción “ni” en combinación con “ya” no tiene el significado de “zelfs”, sino de “zelfs niet meer”.

6. Léxico: otros

En este texto el traductor automático traduce "estúpidos", "tarados" e "idiotas" como "idioten". No sólo estas traducciones no son completamente correctas, sino que también producen una repetición que afecta a la fluidez de la frase.

7. Sintáctico: sujeto erróneo

En la frase 6 el verbo "hoorde" se refiere a un sujeto erróneo "Toen **hij hoorde** dat de bestuurder plotseling remmen" (TF: "Cuando escuchó eso el conductor, freno sorpresivamente"). "Hoorde" se refiere a "hij", mientras que debería referirse a "de bestuurder".

8. Omisión: sujeto

En la frase 5 "los de atrás son unos imbéciles y los de enfrente son unos estúpidos", el traductor automático no reconoce los sujetos "los de atrás" y "los de enfrente". Esta oración es traducida como "de idioten zijn en achterkant zijn dom".

9. Sintáctico: otros

La frase 2 "[...]en el que viajaba mucha gente" es una oración de relativa, que se traduce por otra oración de relativa "die op reis was". La traducción errónea del pronombre relativo tiene influencias en la semántica de toda la frase.

10. Varios

En la primera frase se traduce "érase una vez" por la frase inglesa "once upon a time", para la que existe una traducción neerlandesa adecuada, "er was eens".

En la frase 4 se ha traducido "unos tarados" como "een paar idioten". Sin embargo, en neerlandés lo más idiomático sería "een stel idioten".

4.1.5 Resumen de la evaluación

En primer lugar queremos subrayar que las conclusiones que sacamos aquí no se pueden considerar representativas para todas las traducciones automáticas, sino más bien como un punto de partida que conduce al planteamiento de una hipótesis sobre la influencia del tipo de texto en la calidad de la TA.

- A simple vista parece que la traducción de un texto expresivo (texto 4: el chiste) es más difícil para un traductor automático que la de un texto informativo (texto 1 y texto 3). El texto del *Instituto Cervantes* (texto 1) tiene un error cada 5,7 palabras, el menor número de errores. Después apreciamos que el otro texto informativo, la receta, tiene un error cada 4,5 palabras. El texto turístico (texto 2) tiene un error cada 4,3 palabras y el chiste (texto 4) tiene el mayor número de errores con un error cada 2,9 palabras.
- No parece existir una relación evidente entre el tipo de texto y el tipo de error. Los dos textos informativos no guardan semejanzas respecto a los tipos de errores. El chiste tiene el mayor

número de errores respecto al léxico. Eso se explica por la importancia de la traducción precisa del léxico en los chistes. La broma pierde su sentido por las traducciones de “tarados”, “imbéciles” e “idiotas” como “idioten”. Aunque las palabras “tarados” y “imbéciles” son más o menos equivalentes a “idiotas”, este texto requiere que no se los traduzca como “idiotas”, porque pierde el sentido de la broma. En este sentido podemos concluir que transmitir la forma estética es más difícil para un traductor automático que transmitir el contenido referencial.

- Todos los textos muestran muchos errores con respecto al orden de palabras. Además, hay muchas omisiones en los textos. La omisión del sujeto es producto de la diferencia de las dos lenguas. El español tiene muchas veces el sujeto implícito, mientras que en neerlandés la oración normalmente no puede funcionar sin sujeto explícito.
- Muchos de los errores no se pueden explicar con un análisis lingüístico, ya que no siempre queda claro si un error produce otro error o si es a la inversa. Además, muchos de los errores parecen ser completamente arbitrarios. Como, por ejemplo, la traducción de “revueltos” por “gecodeerde” y las traducciones a otra lengua: “érase una vez” por “once upon a time” y “Buen provecho” por “Bonne appetit”.
- Comparado con los errores léxicos y sintácticos, el traductor automático no tiene muchos problemas con la morfología.

4.2 Ambigüedades en los textos

Como hemos explicado en el capítulo 2, la ambigüedad es uno de los retos más grandes de la TA. Como el traductor automático no es capaz de realizar un “razonamiento de sentido común” e interpretar el contexto, se enfrenta con muchas ambigüedades, que producen muchos errores. Las ambigüedades que abordamos en este apartado difieren de las ambigüedades humanas y, por lo tanto tenemos que “meternos en la cabeza” del traductor automático y olvidar nuestros conocimientos del mundo.

A continuación presentamos los errores producidos por las ambigüedades en los cuatro textos para, después, relacionarlos con las opiniones de los expertos. En otras palabras, comprobamos si hay una relación entre los errores producidos por la ambigüedad y la calidad de las frases traducidas. Muchas veces las oraciones que hemos analizado en el apartado 4.2, tienen, aparte de su error gramatical, una ambigüedad estructural. Por ejemplo, la traducción errónea de una preposición puede producir una ambigüedad estructural.

Evaluaremos las siguientes ambigüedades:

- **Ambigüedad léxica: polisemia u homografía**
- **Ambigüedad léxica categorial**
- **Ambigüedad estructural**
- **Referencias anafóricas¹⁰**

¹⁰ Como hemos explicado en el apartado 2.2, consideramos las referencias anafóricas un tipo de ambigüedad.

4.2.1 La ambigüedad léxica: polisemia u homografía

Como las polisemias y homografías están relacionadas y son difíciles de distinguir, las trataremos de forma conjuntamente en esta evaluación. En la tabla 4.1 presentamos las ambigüedades léxicas y sus respectivas traducciones erróneas y proponemos una traducción alternativa. Por ejemplo, en el primer texto el sustantivo “centro” produce una

ambigüedad. En sí, la traducción de “centro” por “centrum” es correcta, porque produce la misma polisemia. No obstante, en el contexto “centrum” produce una ambigüedad: que puede referirse tanto al centro de una ciudad como al edificio del Instituto. En este texto se refiere al edificio del *Instituto Cervantes* en Utrecht, por tanto proponemos la traducción “*Instituto Cervanto*”.

En el segundo texto encontramos la traducción de “joyas” por “sieraden”. La palabra “joyas” es ambigua en el sentido de que ofrece tanto un significado figurado como uno literal. En este contexto se refiere al significado figurado y, por lo tanto proponemos traducirlo por “juwelen”.

En la frase 7 del tercer texto (T3F7) “freír” es traducido por “frituren”. En español “freír” es según el *DRAE* la acción de “hacer que un alimento crudo llegue a estar en disposición de poderse comer”, así que no incluye los conceptos de las palabras neerlandesas “frituren” y “bakken”. En neerlandés son conceptos distintos; “frituren” se refiere según *Van Dale* a la acción de sumergir el alimento unos minutos en aceite muy caliente. En este texto no se hace referencia al concepto de “frituren”, sino a “bakken”. En la frase 14 el adverbio “poco” produce una ambigüedad. Dependiendo del contexto “poco” puede significar “weinig” o en combinación con el artículo “un” puede significar “een beetje”. Una sartén “poco profunda” se refiere a una sartén “weinig diep” o sea “ondiep”. En cambio, una sartén “un poco profunda” se refiere a una sartén que es “enigszins diep”. En este caso, el traductor automático no ha reconocido el artículo, lo que proporciona una traducción errónea. Probablemente la TA de la receta creará una tortilla diferente a la receta original. En el texto 4 “ni” produce una ambigüedad, dependiendo de su contexto puede tener diferentes significados. La

Frase ¹¹	Texto fuente	Trad. automática	Trad. propuesta
T1F1	Centro	Centrum	<i>Instituto cervantes</i>
T1F5	Entidad	Entiteit	Instelling
T1F10	Acto	gebeurtenis	activiteit
T2F1	Unas	sommige	Enkele
T2F3	Joyas	Sieraden	Juwelen
T2F14		boodschappen	
T2F17	hacer compras	doen	Winkelen
T2F17	naturaleza	Aard	
T3F7	Freír	Frituren	Bakken
T3F12	Ya	Al	eenmaal
T3F14	un poco profunda	Ondiepe	een beetje diepe
T3F17	Lados	Partijen	Kanten
T411	Ya ni	Zelfs	niet meer

Tabla 4.1 Polisemias y homografías

¹¹ Los códigos se refieren al texto (T) y a la frase (F), por ejemplo T1F9 se refiere al texto 1 y a la frase 9

combinación “ya ni”, significa “zelfs niet meer”, así que proponemos la traducción “nu weet ik het zelfs niet meer”.

4.2.2 La ambigüedad léxica categorial

La ambigüedad léxica categorial se produce cuando una palabra se puede referir a más de una categoría gramatical. Además de producir una oración gramaticalmente incorrecta, este tipo de ambigüedad puede acarrear errores pragmáticos en la traducción. En las traducciones automáticas aparecen sustantivos traducidos como adjetivos, verbos traducidos como sustantivos y verbos traducidos como adjetivos.

Sustantivo → Adjetivo

(T1F8) *Mayores de 65: ouder dan 65*

La palabra “mayor” puede ser un sustantivo (“een oudere”) o un adjetivo (“ouder”). Obviamente, en su contexto “mayor” es un sustantivo. La traducción con un adjetivo produce una frase cómica “voor studenten ouder dan 65 jaar”.

(T4F9) *Unos estúpidos: dom*

La palabra “estúpido” puede tener la función de un sustantivo (“een stomkop”) o de un adjetivo (“stom/dom”). Gracias al artículo sabemos que en este contexto se trata de un sustantivo. Sin embargo, es traducido como un adjetivo.

(T4F10) *El borracho: dronken*

Lo mismo ocurre con la palabra “borracho”, que puede tener la función de un sustantivo (“de dronkaard”) o de un adjetivo (“dronken”). El contexto del artículo indica que se trata de un sustantivo, pero es traducido como un adjetivo.

Verbo → Sustantivo

(T2F17) *Vive la naturaleza = de ervaring van de aard*

La palabra “vive” no es una verdadera ambigüedad léxica categorial porque en la lengua fuente sólo puede ser una forma conjugada del verbo “vivir”. Sin embargo, el traductor automático lo ha traducido como un sustantivo, por ello lo mencionamos en esta categoría.

Verbo → Adjetivo

(T1F2) **Organiza debates = georganiseerde debatten**

Lo mismo ocurre con el verbo “organiza”. No ofrece una verdadera ambigüedad léxica categorial, porque en la lengua fuente “organiza” sólo puede ser una forma conjugada del verbo “organizar”. No obstante, de algún modo el traductor automático lo ha traducido como un adjetivo.

(T2F17) **Hecha por dos lados: gemaakt aan twee kanten**

La palabra “hecha” sí ofrece una ambigüedad léxica categorial. Puede ser un participio (“gemaakt”) o un adjetivo (“gaar”). Suponemos que en este caso el autor se refiere al adjetivo.

4.2.3 La ambigüedad estructural

Hablamos de ambigüedad estructural cuando la estructura de una oración se puede interpretar de más de un modo. Distinguimos las ambigüedades estructurales de referencia y las ambigüedades estructurales producidas por una preposición errónea. También hemos incluido las frases que no son ambigüas en su lengua de origen, sino que tienen una ambigüedad en la lengua meta.

1. Ambigüedad estructural de referencia

(T1F4) También ofrece música, teatro, cine y exposiciones **de formato pequeño**,
Ook biedt muziek, theater, film en tentoonstelling van klein formaat,

En esta oración la parte “de formato pequeño” es ambigua, porque puede referirse a “música, teatro, cine y exposiciones”, o sólo a “exposiciones”; sin embargo, no produce un problema, dado que se puede mantener la misma ambigüedad estructural en la TA.

(T2F3) **joyas** del Modernismo y de la arquitectura contemporánea,
de Art Nouveau sieraden en hedendaagse architectuur,

Esta oración no es ambigua para un traductor humano, porque la repetición de la preposición “de” indica que “joyas” tanto se refiere a “Modernismo” como a “la arquitectura contemporánea”. Sin embargo, para el traductor automático sí era ambigua la frase. Relaciona “joyas” (“sieraden”) sólo con “Modernismo” (“Art Nouveau”) y lo cambia a una palabra compuesta.

(T1F6) Las actividades culturales **en el centro suelen tener lugar por la tarde**.
Culturele activiteiten meestal plaats in het centrum in de middag.

Como la oración anterior, esta oración no es ambigua para un traductor humano. El orden de la frase sólo permite una interpretación. Sin embargo, aparentemente esta oración es ambigua para el traductor automático. Puede interpretar esta frase de más de un modo y hace una referencia errónea.

2. Ambigüedad estructural producida por la preposición

Como anteriormente hemos mencionado, las preposiciones dan origen a muchas ambigüedades. La traducción errónea de una preposición puede tener consecuencias para la semántica de una frase. Las ambigüedades se producen de modo especial en las oraciones introducidos por las preposiciones de poca especificidad semántica, es decir, las que son muy frecuentes en español como “a”, “de” y “por”.

(T3F17) Cuando la tortilla está hecha **por** los dos lados, ya se puede comer.

*Wanneer de tortilla wordt gemaakt **door** beide partijen, kan worden gegeten.*

En esta oración “por” no tiene el significado de “door”, sino de “aan”. La oración cambia totalmente por la traducción errónea de la preposición. Probablemente la traducción errónea de la preposición está relacionado con la interpretación errónea de 'hecha' como participio y la traducción de "lado" como "partij".

(T3F7) Para eso hay que freirlas **a** fuego bajo

Dit wordt gedaan door frituren warmte laag

Debido al hecho de que la preposición “a” aparece muchas veces en español donde no aparece en neerlandés, el traductor automático ha omitido esta preposición en la traducción. Sin embargo, en esta oración “a” se traduce con “op”.

(T3F4) mercados **para** regalarse los sentidos, tesoros de la antigua ciudad romana y medieval,

*de markten **voor** de zintuigen genieten van de schatten van de oude Romeinse en middeleeuwse stad,*

En esta frase se traduce “para” como “voor”, con la consecuencia de que las relaciones dentro de la frase cambian. En este frase “para” aparece en combinación con un infinitivo y, por lo tanto, tiene otro significado. Introduce un fin o un objetivo, lo que se puede traducir como “om te”.

4.2.4 Referencias anafóricas¹²

Una anáfora se refiere a una unidad mencionada de forma explícita en otro lugar del texto. Si no queda claro a qué antecedente se refiere la anáfora, hablamos de una ambigüedad. Como el traductor automático no puede interpretar el contexto, podemos imaginarnos que es un problema considerable. Presentamos algunos ejemplos en los que el traductor automático se refiere al antecedente equivocado. Lo que llama la atención es que en la mayoría de los casos el traductor automático ignora las anáforas.

(T3F7) Para eso hay que freír^{las} a fuego bajo
Dit wordt gedaan door frituren warmte laag

En esta oración el antecedente de “las” es “patatas” (T3F6 “[...] es importante que las patatas estén blandas[.]”). El traductor automático ignora la anáfora “las”, lo cual produce una frase incomprensible.

(T4F5) ^{los} de atrás son unos imbéciles y ^{los} de enfrente son unos estúpidos.
de idioten zijn en achterkant zijn dom.

El traductor automático tiene mucha dificultad con la traducción de esta frase. No reconoce la anáfora “los” y, además, omite la parte “de enfrente”. Esta traducción es totalmente incomprensible.

4.2.5 Resumen de la evaluación

Debido a las limitaciones de los ordenadores, la ambigüedad es uno de los retos más grandes de la TA. En el lenguaje humano hay muy pocas oraciones que realmente sean ambiguas, porque en la mayoría de los casos podemos deducir la interpretación exacta del contexto o de nuestro conocimiento del mundo. Curiosamente, las ambigüedades a las que se enfrenta el traductor automático se ponen de manifiesto a la hora de una traducción errónea. En la mayoría de las oraciones no vemos ninguna ambigüedad hasta que el traductor automático traduzca el texto. Por tanto, sólo hemos discutido las ambigüedades que se ponen de manifiesto por una traducción automática errónea. Podemos sacar las siguientes conclusiones:

- La ambigüedad puede producir un texto completamente incomprensible, o peor, puede procurar que el lector se incline por una interpretación errónea. Otros errores lingüísticos pueden producir frases incorrectas, pero los errores causados por la ambigüedad pueden producir una frase muy

¹² Como en los textos casi sólo aparecen referencias anafóricas, hemos decidido ignorar las referencias catafóricas. Véase el apartado 2.2.3

distinta a la original. Por ejemplo, la traducción de "lados" por "partijen" (T3F17) produce una oración completamente incomprensible.

- Las ambigüedades aparecen en todos los textos, no hay una relación evidente entre el tipo de texto y la ambigüedad. Con otras palabras, los errores causados por la ambigüedad son inherentes a la TA.
- La ambigüedad léxica polisémica u homográfica es el tipo de ambigüedad que aparece más veces. A primera vista, este tipo de ambigüedad parece tener una gran influencia en la semántica de una oración. Debido a la ambigüedad léxica, la TA de, por ejemplo, la receta difiere en unos puntos esenciales del original.
- La ambigüedad léxica categorial sólo aparece en algunas oraciones. Los sustantivos traducidos por adjetivos son más frecuentes.
- Las preposiciones desempeñan un papel especial en la ambigüedad estructural. Las preposiciones de poca especificidad semántica como "a", "de" y "por" dan origen a varias ambigüedades.
- Las referencias anafóricas las consideramos como un tipo de ambigüedad, porque el antecedente al que se refiere la anáfora, puede ser ambiguo. Lo que llama la atención es que en estos textos, el traductor automático muchas veces no es capaz de identificar la anáfora. Por lo tanto, no es la ambigüedad la que produce los errores, sino la incapacidad del traductor automático de identificar las anáforas.

4.3 Las valoraciones de los expertos

En los apartados anteriores hemos evaluado los textos en detalle. Hemos identificado los errores lingüísticos y los errores producidos por la ambigüedad. Como un "FAHQT"¹³ está fuera del alcance actual y los errores probablemente siempre estarán involucrados en la TA, queremos localizar la importancia de cada error y su influencia en la calidad de la oración.

Puede ser que un error concreto no sea tan grave para la calidad de una traducción. Usamos las valoraciones de los expertos para medir la calidad de la traducción. La opinión del experto es interesante, porque, aparte de su pericia, es la opinión del usuario potencial. Es decir, las valoraciones de los expertos son de importancia por su fiabilidad y su relevancia. Por lo general, podemos concluir que los expertos no son muy optimistas acerca de las traducciones automáticas. Un traductor profesional dijo "los resultados varían de hilarante a lamentable", otro dijo "no tengo la impresión de que mi medio de vida esté en peligro", y otro traductor profesional consideraba la TA "peligrosa", porque dijo "Creo que es bastante peligroso traducir con este tipo de programas. Las frases ofrecidas pueden hacer que te inclines por una interpretación errónea". También había comentarios más positivos. Un traductor profesional indicaba que a pesar de que la TA no es perfecta, sí puede servir como base de una traducción.

¹³ Fully automatic high quality translation.

Las valoraciones medias de los expertos nos permitirán clasificar las oraciones y servirán de base para el análisis de la relación entre los errores lingüísticos, la ambigüedad y la calidad de la TA. Obviamente los expertos no evalúan de la misma manera, así que primero queremos presentar las valoraciones de cada evaluador individual. Cuando examinamos estas valoraciones de cada evaluador individual observamos diferencias. La tabla 4.3.1 presenta la media aritmética más alta y la más baja de los evaluadores individuales y la diferencia entre los dos valores.

	mas alto	mas bajo	Diferencia
texto 1	3,70	1,50	2,20
texto 2	3,35	1,98	1,37
texto 3	3,06	1,61	1,45
texto 4	2,18	1,00	1,18

Tabla 4.2 Diferencia entre el texto mejor y el peor evaluado

Por ejemplo, los valores del cuarto texto están entre 1,00 y 2,18, lo que quiere decir que el valor medio más alto del evaluador individual, o sea el evaluador menos severo, es 2,18 y el del evaluador más severo es 1,00. Es decir, hay una diferencia de 1,18 entre las evaluaciones más altas y las más bajas de este texto. Como indica la tabla, las diferencias entre el mayor y el menor valor son aún más altas en los otros textos, hasta 2,20 en el primer texto. No obstante, cuando analizamos los resultados más en detalle descubrimos que los evaluadores que asignan valoraciones bajas a un texto, tienden a asignar valoraciones bajas a todos los textos. Es la misma persona la que da las evaluaciones más altas (un estudiante de traducción), y la misma persona la que da las evaluaciones más bajas (un traductor profesional). Esto sugiere que los evaluadores aplican diferentes normas para evaluar la TA, y que en todo el cuerpo de los textos sus normas tienden a permanecer relativamente constantes. Por lo general, los traductores profesionales son más negativos frente a las traducciones automáticas que los estudiantes del Máster de Traducción. En los siguientes apartados no analizamos las valoraciones del evaluador individual, sino las valoraciones medias.

4.3.1 El tipo de texto y la calidad

Las oraciones valoradas por los expertos pertenecen a una determinada tipología textual basada en la función comunicativa del texto, a saber informativo, operativo o expresivo. La tabla 4.3 presenta las valoraciones medias de los evaluadores por tipo de texto. Es decir, el promedio de las oraciones valoradas en una escala de 1 (mala calidad) a 5 (buena calidad) ordenado por tipo de texto. De esta tabla se desprende que, excepto en el texto expresivo, no hay una diferencia evidente entre los valores de los textos. El texto operativo ha recibido el valor más alto y el texto expresivo el valor más bajo. Los textos informativos tienen aproximadamente el mismo valor. Esta tabla confirma lo que hemos concluido de la evaluación cualitativa, que traducir automáticamente un chiste es más difícil que traducir un texto turístico, una receta o un texto informativo. Por lo tanto, podemos sacar la conclusión, no totalmente inesperada, de que “crear un efecto de equivalencia” es más fácil para un

traductor automático que “transmitir la forma estética”, y “representar hechos” es más fácil que “adoptar la perspectiva del autor del texto fuente”.

Texto		Topología	Media
Texto 1	cervantes	Informativo	2,34
Texto 2	Turístico	Operativo	2,76
Texto 3	Receta	Informativo	2,33
Texto 4	Chiste	Expresivo	1,68

Tabla 4.3 Valoraciones medias por tipología textual

La tabla 4.4. presenta las oraciones mejor y peor evaluadas en cada texto y la diferencia entre ambas. La frase mejor evaluada en el texto 2, por ejemplo, tiene un valoración de 4,88, mientras que la frase peor evaluada en texto 2 tiene una valoración de 1,00. Por lo tanto, podemos concluir que la calidad de este texto no es constante. Lo que llama la atención es que los textos con una valoración baja (véase la tabla 4.3), suelen mostrar una diferencia baja entre la mejor y la peor oración. En otras palabras, todas las oraciones individuales tienen una valoración baja en una TA de relativamente mala calidad, es decir que sólo raramente las oraciones individuales tienen una valoración alta. Al contrario, en las traducciones automáticas con una calidad relativamente alta sólo hay algunas oraciones que tienen una valoración baja.

Texto	Mejor oración	Peor oración	Diferencia
texto 1	3,63	1,50	2,13
texto 2	4,88	1,00	3,88
texto 3	4,86	1,00	3,86
texto 4	2,57	1,00	1,57

Tabla 4.4 Las oraciones mejor y peor evaluadas.

Obviamente, nos interesa saber qué elementos hacen que los expertos valoren bajas o altas ciertas oraciones, es decir, por qué la intuición lingüística del evaluador califica como mala o buena a una oración. En la tabla 4.5 presentamos las seis mejores oraciones y las seis peores oraciones con las oraciones originales. A pesar de que las mejores oraciones son gramaticalmente correctas, fluidas y fieles al texto fuente, ninguna oración ha recibido un 5 como valor medio. Es decir, según la opinión media de los expertos, ninguna de las oraciones está perfecta. De las oraciones mejor evaluadas cuatro pertenecen al texto turístico y dos a la receta. Las peores oraciones son casi completamente incomprensibles o presentan un contenido erróneo. Todos los evaluadores han asignado el valor más bajo a estas oraciones. Lo que caracteriza a las peores oraciones es que aparte de que no son gramaticales, los errores influyen en la semántica de la frase. Por ejemplo, la omisión de la negación en T2F7, la traducción errónea de “lados” (T3F17), “ya ni” y “revueltos” (T4F11), y las referencias anafóricas erróneas en T4F2 y T4F5 tienen una gran influencia en la

fidelidad, comprensibilidad y fluidez de las oraciones. De las oraciones peor evaluadas cuatro aparecen en el chiste, una en el texto turístico y una en la receta.

Fuente	Mejores oraciones	Valor
T3F3	(se calcula una patata grande por persona) (naar schatting een grote aardappel per persoon).	3,83
T2F16	Y descubre la gastronomía de temporada En ontdek de keuken van het seizoen,	4
T2F9	Barcelona te ofrece opciones interesantísimas para todas las épocas del año. Barcelona biedt interessante mogelijkheden voor alle seizoenen.	4,13
T2F8	Esta primavera en Barcelona Dit voorjaar in Barcelona	4,63
T3F2	Se pelan y se lavan las patatas Schil en was de aardappelen	4,86
T2F13	Pero hay muchísimo más. Maar er is zoveel meer.	4,88
Fuente	Peores oraciones	Valor
T2F7	Y lo mejor es que no hay que ser un gran explorador para encontrarlos y conocer todos los barrios de Barcelona. En het beste deel is dat er een grote ontdekkingsreiziger te vinden en kennen de buurten van Barcelona.	1,00
T3F17	Cuando la tortilla está hecha por los dos lados, ya se puede comer. Wanneer de tortilla wordt gemaakt door beide partijen, kan worden gegeten.	1,00
T4F11	Ya ni sé, ¡Están todos revueltos! Ik weet zelfs niet weet, worden alle gecodeerde!	1,00
T4F2	un borracho que abordó un autobús en el que viajaba mucha gente, een dronkelap die aan boord van een bus die op reis was een heleboel mensen,	1,00
T4F5	los de atrás son unos imbéciles y los de enfrente son unos estúpidos. de idioten zijn en achterkant zijn dom.	1,00
T4F6	Cuando escuchó eso el conductor, freno sorprendentemente, Toen hij hoorde dat de bestuurder plotseling remmen,	1,00

Tabla 4.5 Las oraciones mejor y peor evaluadas

Cuando observamos la tabla 4.4 no es muy difícil pensar en por qué algunas oraciones están mejor calificadas que otras. Nuestra intuición lingüística nos dice que algunas oraciones son relativamente buenas y otras relativamente malas. Lo que produce exactamente esta calificación, es lo que vamos a estudiar más en detalle en el siguiente apartado. Compararemos las valoraciones con los errores lingüísticos y los errores producidos por la ambigüedad.

4.3.2 Los errores y la calidad

En una pregunta abierta los evaluadores podían indicar qué tipo de error les molestaba más. Casi cada experto indicó que los errores pragmáticos le molestaba más. Un traductor profesional dijo: "Un error gramatical es feo pero una interpretación errónea puede acarrear problemas." También los errores morfológicos molestaron a los evaluadores, "Muchas veces el sujeto no concuerda con la forma verbal" y "El traductor hace combinaciones erróneas relacionando el verbo con el sujeto equivocado." Las omisiones también eran un problema frecuentemente mencionado. En la tabla 4.5 hemos clasificado los errores que hemos evaluado en el apartado 4.1 según la calidad de las oraciones en las que aparece el error. Denominamos "mala calidad" a las oraciones con una valoración hasta 2,00, "calidad media" a las oraciones con una valoración entre 2,00 y 3,50 y "buena calidad" a las oraciones con una valoración entre 3,50 y 5,00. Por ejemplo, el error léxico "palabra errónea" aparece 14 veces en las oraciones de mala calidad y 2 veces en las oraciones de buena calidad. Hemos encontrado un total de 143 errores en los textos, de los que 103 aparecen en las oraciones de mala calidad, 35 en las oraciones de calidad media y 5 en las oraciones de buena calidad. Según cabía esperar, estos valores confirman que el número de errores en cada oración es el factor determinante en la evaluación de los expertos. Aunque los expertos usaban su intuición lingüística para evaluar las oraciones y no identificaban los errores individuales, valoraban más bajo las oraciones con un gran número de errores.

	mala	media	Buena	Total
léxico: palabra errónea	14	2	0	16
léxico: ambiguo	11	5	0	16
léxico: otros	4	5	1	10
función gramatical errónea	7	5	0	12
sintáctico: sujeto erróneo	5	1	0	6
sintáctico: orden de palabras	15	4	0	19
sintáctico: otros	6	0	0	6
morfológico: conjugación errónea	11	3	0	14
omisión: verbo	4	1	0	5
omisión: sujeto	8	0	0	8
omisión: otros	8	2	2	12
preposición errónea	6	3	1	10
Varios	4	4	1	9
Total	103	35	5	143

Tabla 4.6 Errores según tipo de error y calidad de las frase.

La mayoría de los errores que hemos encontrado en los textos son errores léxicos, un total de 42 errores. En segundo lugar tenemos los errores sintácticos con un total de 31 errores, de los que

19 son errores con respecto al orden de palabras. Otro tipo de error que se produce bastantes veces es la omisión de palabras, que aparece 25 veces. No obstante, las oraciones no necesitan ser sin errores para recibir la calificación “buena calidad”. Aparte del número de errores en una oración, el tipo de error también puede influir en la evaluación de la calidad de una oración traducida automáticamente. Sin embargo, es difícil concluir de esta tabla la causa exacta de la valoración baja o alta. Puede ser el tipo de error o la cantidad de errores que causan una valoración baja. Cuando observamos las oraciones individualmente podemos ver que los expertos valoran muy bajo las traducciones a otras lenguas como “¡Buen provecho!” por “Bonne Appetit!” (3F18) y la traducción de “Érase una vez” por “Once upon a time” (T4F1). Es decir, los expertos evalúan la traducción a otra lengua peor que, por ejemplo, una preposición errónea. Es muy complejo explicar por qué el traductor automático hace ciertos errores. El gran número de errores léxicos no se deja explicar fácilmente. Como hemos mencionado anteriormente, las traducciones de algunas palabras parecen totalmente arbitrarias, como por ejemplo “revueltos” por “gecodeerde”. Este tipo de errores se debe probablemente a la arbitrariedad del método estadístico del traductor automático. Otros errores léxicos se explican por las dificultades intrínsecas a la ambigüedad. El gran número de omisiones es llamativo y probablemente el resultado del método estadístico del *Traductor Google*.

4.3.3 La ambigüedad y la calidad

Como los expertos indicaban en las preguntas abiertas, los errores pragmáticos les molestan más, y como a veces son el resultado de la ambigüedad, hemos analizado la relación entre la calidad y los errores producidos por la ambigüedad. La tabla 4.6 presenta las ambigüedades en los textos y la calidad de las oraciones. Un total de 37 errores en los textos está relacionado con la ambigüedad. Muchos de estos errores aparecen en las oraciones de mala calidad. Para ser más precisos, un 68 % de todas las oraciones de mala calidad tiene un error relacionada con la ambigüedad. En las oraciones de buena calidad no encontramos errores producidos por la ambigüedad. Nos podemos preguntar si una oración no puede ser buena si contiene un error producido por la ambigüedad. Lo que llama la atención es que el número de ambigüedades en los diferentes textos es aproximadamente el mismo. Es decir que la ambigüedad aparece en cada tipo de texto, no importa si es un texto informativo, operativo o expresivo.

	mala	Media	buena	Total
ambigüedad léxica	10	5	0	15
ambigüedad categorial	3	5	0	8
ambigüedad estructural	7	2	0	9
referencia anafórica	5	0	0	5
Total	25	12	0	37

Tabla 4.7 Errores según ambigüedad y calidad de las frase.

Los errores producidos por la ambigüedad léxica son los más frecuentes; 10 de estos errores se encuentran en las oraciones de mala calidad y 5 en las oraciones de calidad media. Las ambigüedades léxicas que encontramos en las oraciones de mala calidad son, por ejemplo, “freír” por “frituren” y “lados” por “partijen”. En las oraciones de calidad media encontramos errores menos “graves” como por ejemplo la traducción de “centro” por “centrum”. Un ejemplo de las ambigüedades léxicas categoriales que encontramos en las oraciones de mala calidad es el sustantivo “mayores”, traducido por el adjetivo “ouder”, lo que cambia la oración en “met een korting van \$2 voor studenten ouder dan 65 [...]”. La traducción del sustantivo “borracho” por el adjetivo “dronken” se encuentra en una oración calificada como de calidad media: “en heel erg boos de bestuurder nam de dronken bij de kraag”. Todas las referencias anafóricas erróneas se encuentran en las oraciones de mala calidad.

4.3.4 Resumen de la evaluación

Para medir la calidad de una TA solamente clasificar los errores no es suficiente. Puesto que los errores probablemente siempre estarán involucrados en la TA, necesitamos otro parámetro para decir algo sobre la calidad. Opinamos que las valoraciones de los expertos son esenciales, porque nos permiten calificar las oraciones con sus errores, lo cual nos proporciona las siguientes conclusiones:

- Por lo general, los expertos no son muy optimistas frente a la TA. Los traductores profesionales eran más pesimistas que los estudiantes del Máster de Traducción. Ninguna oración es valorada con un 5 como valor medio, lo que significa que aparentemente los evaluadores tienen grandes expectativas de lo que es una TA perfecta. Sin embargo, también es posible que tenga que ver con la actitud negativa frente a la TA.
- Los expertos valoran el chiste más bajo, porque todas las oraciones en este texto tienen una valoración baja. Los otros textos tienen aproximadamente la misma valoración. El texto turístico tiene valoración más alta, pero no hay una diferencia considerable con los textos informativos. Este resultado podemos considerar como una conclusión provisional de que es más difícil traducir automáticamente un texto expresivo que un texto informativo u operativo.
- El número de errores en una oración es el factor determinante en la evaluación de los expertos. Además del número de errores, el tipo de errores puede influir en la calidad de una TA. Una palabra o expresión traducida a otra lengua que la lengua meta, es valorada muy baja.
- La mayoría de los errores en el texto son errores léxicos y sintácticos de los que la mayoría pertenece a un orden de palabras erróneo. El gran número de palabras omitidas también es llamativo. Es difícil explicar por qué el traductor automático hace unos errores más que otros. A veces la traducción parece totalmente arbitraria, lo cual se debe probablemente a la arbitrariedad del método estadístico del traductor automático.
- Muchos errores se originan en la ambigüedad, que es inherente a una lengua y aparece en cada tipo de texto.
- La ambigüedad parece ser una de las razones principales para calificar una oración como mala. La mayoría de las oraciones de mala calidad contiene un error producido por la ambigüedad, que,

a diferencia de otros errores lingüísticos, puede producir un texto completamente incomprensible o presentar un contenido erróneo.

- Los errores producidos por la ambigüedad léxica son los más frecuentes.

4.4 La calidad después de ‘post-editar’

En los apartados anteriores hemos medido la calidad de la TA mediante un análisis de errores y mediante valoraciones de expertos. A través de la post-edición sin texto fuente intentamos sacar conclusiones sobre la calidad de las traducciones automáticas desde otro punto de vista: analizamos la posibilidad de producir un texto correcto mediante la post-edición. Nos preguntamos si los errores en una TA realmente dicen algo sobre la calidad si un revisor puede corregir fácilmente estos errores. En otras palabras, ¿no es la facilidad de corregir una TA lo que realmente valora la calidad?

La post-edición es la revisión y corrección de un texto traducido. Es un proceso que también se realiza en las traducciones humanas. Evidentemente, las revisiones de una traducción automática y de una traducción humana son muy distintas. Los errores que se encuentran en una TA difieren mucho de los errores en una traducción humana. Un traductor humano puede cometer accidentalmente un error, mientras que en la mayoría de los casos un traductor automático comete el mismo error repetidamente. Además, hemos visto que el traductor automático puede traducir una palabra ambigua completamente errónea y, por consiguiente, ofrecer una interpretación errónea. Según Nirenburg (1987) la post-edición de una TA puede tener un efecto contraproducente:

Translation quality may be low on stylistic or semantic grounds. The decision to produce ‘rough and dirty’ translations by passing semantical analysis can be misleading. Consider a cookie jar with 100 delicious freshly baked chocolate-chip cookies, 95 of which are not poisoned. Such a cookie jar could be worse than useless; it could be tempting and therefore actively dangerous.

Eso es particularmente el caso de las traducciones automáticas de mala calidad. Esperamos que la revisión del chiste (con un valor medio de 1,68) será un buen ejemplo de un ‘cookie jar with poisoned cookies’. Revisar este texto probablemente será poco eficaz. Hay tantos errores en este texto que creemos que sería más eficaz traducir este texto de nuevo.

4.4.1 La post-edición sin texto fuente

Para extraer los errores difíciles de revisar, dos traductores humanos han revisado la TA del texto del *Instituto Cervantes* sin texto fuente. Se pueden encontrar las dos traducciones post-editadas en los anexos 5 y 6. Evidentemente, estamos particularmente interesados en cómo los revisores han corregido los errores de la TA. Comparamos los errores antes de la post-edición con los errores después de la post-edición. Llamamos “errores fáciles de revisar” a los errores corregidos correctamente por los revisores, y llamamos “errores difíciles de revisar” a los errores que todavía están presentes en el texto o no están corregidos correctamente. Además, nos interesa analizar cómo

los revisores tratan los errores producidos por la ambigüedad. Ambos revisores han indicado que en la vida real nunca aceptarían tal trabajo de traducción sin texto fuente y que revisar este texto requiere mucho adivinar y que es muy probable que estén influidos por sus conocimientos anteriores del *Instituto Cervantes*. Además de la revisión de cada oración, presentamos la valoración de los expertos sobre la calidad de la TA de cada oración. De esta manera podemos observar si hay una relación entre la facilidad de corregir una oración y la calidad de la oración.

1. Errores fáciles de revisar

Errores léxicos: palabra ambigua

Calidad 2,25

TF: El **centro** de Utrecht organiza debates [...]

TA: Het **centrum** van Utrecht organiseert debatten [...]

PE1: Ons **centrum** in Utrecht organiseert debatten [...]

PE2: Het **Instituto Cervantes** in Utrecht organiseert debatten [...]

En este contexto “centro” es ambiguo porque puede referirse tanto al centro de una ciudad como a un edificio. Los revisores lo han solucionado añadiendo “ons” o reemplazándolo por *Instituto Cervantes*, y reemplazando la preposición “de” por “in”.

Errores léxicos: palabra errónea

Calidad 1,5

TF: **siempre** con traducción simultánea español-neerlandés y a la inversa.

TA: **mits** met gelijktijdige Spaans-Nederlands en vice versa

PE1: **met** simultaanvertaling Spaans-Nederlands en vice versa.

PE2: **steeds** met simultaanvertaling Spaans-Nederlands en vice versa.

La TA de “siempre” es totalmente errónea. No obstante, los revisores corrigen la oración relativamente correctamente, probablemente gracias a sus conocimientos del *Instituto Cervantes*.

Omisión: sujeto

Calidad 1,75

TF: También ofrece música [...]

TA: Ook biedt muziek [...]

PE1: Ook bieden **wij** muziek [...]

PE2: Ook biedt **het** muziek [...]

Aparentemente la omisión del sujeto no es un gran problema para los revisores. En la primera post-edición el revisor ha cambiado la perspectiva del texto, en la segunda post-edición el revisor ha añadido el sujeto ‘het’.

Errores léxicos: otros

Calidad 2,5

TF: **tanto** en el mismo centro **como** [...]

TA: **zowel** in het centrum **en** [...]

PE1: **zowel** in het eigen gebouw **als** [...]

PE2: **zowel** in het eigen centrum **als** [...]

La TA de “tanto .. como” como “zowel .. en” no es correcta. La colocación fija es “zowel .. als”, un tipo de error que los revisores pueden corregir fácilmente.

Omisión: verbo

Calidad 1,63

TF: Las actividades culturales en el centro **suelen** tener lugar por la tarde.

TA: Culturele activiteiten meestal plaats in het centrum in de middag.

PE1: De culturele activiteiten **vinden** meestal ‘s middags in het centrum plaats.

PE2: De culturele activiteiten in ons centrum **vinden** over het algemeen in de middag plaats.

A pesar de que los expertos han valorado muy bajo esta oración, los revisores pueden corregirla fácilmente. La omisión del verbo no es un problema para los editores, el contexto revela que lo que falta aquí es “vinden”.

Errores sintácticos: orden de palabras

Calidad 1,63

TF: Habitualmente, el **público asistente tiene la ocasión** de formular preguntas al término del acto.

TA: Meestal is **het publiek heeft de kans** om vragen te stellen na de gebeurtenis.

PE1: Meestal **krijgt het publiek de gelegenheid** om vragen te stellen na de bijeenkomst.

PE2: Na afloop van elke activiteit **is er vaak gelegenheid** om vragen te stellen.

Esta oración también tiene una valoración muy baja. Sin embargo, los revisores solucionan el orden de palabras errónea fácilmente.

Varios

Calidad 3,13

TF: o **escribiendo** a nuestro departamento cultural: cultutr@cervantes.es

TA: of **schrijven** naar onze culturele afdeling: cultutr@cervantes.es

PE1: of **een mail te sturen** aan onze culturele afdeling: cultutr@cervantes.es

PE2: of **via een bericht** aan onze culturele afdeling: cultutr@cervantes.es

Lo que falta en la TA es el “te” antes del infinitivo “schrijven”. Los revisores han solucionado este problema cambiándolo por “een mail te sturen” o “via een bericht”.

2. Errores difíciles de revisar

Ambigüedad léxica categorial

Calidad 1,13

TF: con un descuento de 2 euros para **estudiantes, mayores de 65** y socios de la biblioteca

TA: met een korting van \$ 2 voor **studenten ouder dan 65** en de leden van de bibliotheek

PE1: met een korting van 2 euro voor **mensen ouder dan 65** en voor leden van de bibliotheek

PE2: met een korting van 2 euro voor **studenten, 65-plussers** en bibliotheek leden.

El traductor automático ha traducido el sustantivo “mayores” con el adjetivo “ouder”, lo cual cambia el significado de la frase. En la primera post-edición el revisor ha corregido falsamente “studenten” por “mensen”. El otro revisor ha interpretado (o adivinado) el significado adecuado.

Ambigüedad estructural

Calidad 1,63

TF: **Las actividades culturales en el centro** suelen tener lugar por la tarde.

TA: **Culturele activiteiten** meestal plaats **in het centrum** in de middag.

PE1: **De culturele activiteiten** vinden meestal 's middags **in ons het centrum** plaats.

PE2: **De culturele activiteiten in ons centrum** vinden over het algemeen in de middag plaats.

En esta oración “en el centro” se refiere a “las actividades culturales”, mientras que en la TA se refiere a “meestal plaats”. En la primera post-edición se mantiene esta referencia errónea, mientras que en la segunda es corregida correctamente.

Errores léxicos: otros

Calidad 2,5

TF: Los interesados pueden recibir un programa trimestral en papel y/o electrónico **con el recordatorio** de cada actividad.

TA: Wie geïnteresseerd is kan ontvangen ieder kwartaal een papier en/of elektronische nieuwsbrief **met een herinnering** van elke activiteit.

PE1: Belangstellenden kunnen ieder kwartaal een papieren of elektronische nieuwsbrief ontvangen **met informatie over de activiteiten**.

PE2: Belangstellenden ontvangen elk kwartaal een papieren en / of elektronische nieuwsbrief **plus een uitnodiging voor elke activiteit**

Aunque los expertos no han valorado muy bajo la calidad de esta oración, resulta bastante difícil para los revisores ofrecer una traducción correcta. En este caso los revisores han interpretado “met een herinnering” libremente. En la primera post-edición el revisor ha corregido “herinnering” por “informatie” y en la segunda post-edición el revisor lo ha corregido por “uitnodiging”. Además, ha cambiado “met” por “plus”. Ninguna de estas correcciones es fiel al texto fuente.

4.4.2 Resumen de la evaluación

A la hora de evaluar la calidad de una TA es muy importante, aparte del producto crudo, tener en cuenta el producto post-editado. Hemos visto que muchos errores que en primera instancia parecen errores 'graves', son fáciles de corregir. Después del análisis tenemos que sacar las siguientes conclusiones:

- Cuando comparamos las valoraciones de los expertos, o sea la calidad de la frase, con la facilidad de corregir una oración, no vemos una relación. Es decir, las oraciones con una valoración baja, no siempre son las más difíciles de revisar.
- Los revisores pudieron corregir la mayoría de los errores en la TA sin tener el texto fuente. Había sólo algunos errores que el revisor no podía corregir, pero esperamos que se podrán corregir con el texto fuente.
- Pensamos que los “conocimientos del mundo” (mejor dicho, los conocimientos del *Instituto Cervantes*) juegan un papel muy importante a la hora de revisar este texto. Gracias a estos conocimientos los revisores podían “adivinar” la oración correcta.
- Los revisores tienen poca dificultad con la corrección de errores “superficiales”, es decir, los errores visibles, como por ejemplo, un orden de palabras erróneo, errores léxicos, etcétera. El gran peligro se encuentra en las oraciones aparentemente correctas, o sea, las oraciones con errores invisibles. Con estas oraciones uno corre el riesgo de no reconocer los errores.

4.5 Conclusión

Para presentar la calidad de la TA lo más completa posible, hemos abordado la calidad desde diferentes perspectivas. Hemos evaluado lingüísticamente un texto sobre las actividades culturales del *Instituto Cervantes*, un texto turístico sobre Barcelona, una receta de tortilla de patatas y un chiste. El texto sobre el *Instituto Cervantes* tiene, en proporción, el menor número de errores y el chiste el mayor número de ellos. A pesar de que hemos comentado en el capítulo 2 que la consistencia es una ventaja de la TA, en realidad resulta que el traductor automático es muy arbitrario. Omite y añade palabras, hace “adivinations salvajes” y sus errores no son consistentes. La mayoría de los errores en los cuatro textos son errores léxicos, seguidos de errores sintácticos, de los que la mayoría son errores con respecto al orden de palabras. El número de errores en una oración es el factor determinante en la evaluación de los expertos. Las oraciones que contienen un gran número de errores, muchas veces tienen una valoración baja. Además del número de errores en una oración, el tipo de error influye en la calidad de una TA. La mayoría de las oraciones de mala calidad contiene un error producido por la ambigüedad, que puede producir un texto completamente incomprensible o presentar un contenido erróneo, más habitualmente que otros errores lingüísticos. Los errores producidos por la ambigüedad parecen ser una de las razones principales para calificar una oración como mala.

Como tanto una traducción humana como una traducción automática siempre va a ser revisada, hemos medido la calidad desde otro punto de vista. Nos preguntamos si el número o tipo de errores cuenta si el revisor puede corregirlos fácilmente. La post-edición de la TA del texto del *Instituto Cervantes* produjo un texto de calidad razonable. Las oraciones con una valoración baja no fueron siempre las oraciones más difíciles de revisar. Los errores más difíciles fueron los errores “invisibles”, y son éstos los que muestran el peligro de post-editar una TA sin texto fuente.

Conclusión

Como la traducción automática se incorpora cada vez más en el trabajo de los traductores, nos pareció interesante entender mejor esta aplicación y su producto. En este estudio hemos intentado contestar a la pregunta *¿Cuál es el alcance actual de la traducción automática en cuanto al par de lenguas español-neerlandés?*

Contestar a esta pregunta fue difícil, porque ¿cómo se mide la calidad de una traducción automática? Opinamos que sólo contar errores lingüísticos no ofrece una respuesta suficientemente específica a la pregunta de estudio. Por eso, creemos que para presentar la calidad de la TA lo más completo posible, debemos analizar los textos traducidos automáticamente desde diferentes perspectivas. Primero hemos evaluado lingüísticamente cuatro textos con diferentes funciones comunicativas, después un grupo de expertos ha valorado estos textos con la ayuda de su intuición lingüística y, por fin, dos traductores profesionales han revisado un texto traducido automáticamente sin texto fuente. En este estudio queríamos dar atención especial a la ambigüedad, un fenómeno que según muchos es el reto más grande de la TA. Hemos obtenido los siguientes resultados interesantes, que se pueden considerar como punto de partida para iniciar más investigaciones.

Primero concluimos que hay una diferencia entre la calidad de la traducción de los diferentes tipos de textos. El *Traductor Google* parece ser más adecuado para la traducción de textos informativos y operativos y menos adecuado para la traducción de textos expresivos. Eso no sólo se desprende de la evaluación lingüística, sino también de las opiniones de los expertos. En otras palabras, transmitir la forma estética es más difícil para un traductor automático que transmitir el contenido.

La ambigüedad desempeña un papel significativo en el análisis de los errores lingüísticos: muchos de los errores se originan en la ambigüedad. Curiosamente, las ambigüedades a las que se

enfrenta el traductor automático sólo se ponen de manifiesto a la hora de que aparezca una traducción errónea. Esto es porque muy pocas oraciones son realmente ambiguas para los traductores humanos, mientras que para el traductor automático, que no puede deducir la interpretación del contexto o de su conocimiento del mundo, muchas oraciones y palabras sí son ambiguas. Por tanto, sólo vemos las ambigüedades con una traducción errónea. Más que otros errores, la ambigüedad puede producir un texto completamente incomprensible, o peor, puede presentar un contenido erróneo. Esto probablemente explica la valoración baja de las oraciones que contienen un error producido por la ambigüedad.

Como tanto una traducción humana como una traducción automática siempre serán revisadas, hemos medido la calidad mediante la post-edición sin texto fuente. Lo que llamaba la atención era que las oraciones con una valoración baja no siempre eran las más difíciles de revisar. Por lo general, los revisores podían corregir los errores en el texto sin tener el texto fuente. Había sólo algunos errores que no podían corregir de esta manera.

La TA no ha llegado a ser lo que algunos investigadores han previsto; no existe un *fully automatic high quality translation* y la profesión del traductor no ha desaparecido. No obstante, y teniendo en cuenta el tipo de texto, los experimentos han mostrado que la TA puede producir un texto de calidad razonable. Por tanto, prevemos que la incorporación de la TA cambiará el trabajo de los traductores en el sentido de que éstos serán cada vez más revisores.

Referencias

- ALPAC, 1966, *Languages and machines: computers in translation and linguistics*. Washington, D.C., National Academy of Sciences, National Research Council.
- Arnold, D., 2003, "Why translation is difficult for computers"., En: Somers, H., *Computers and Translation: A translator's guide*, Amsterdam, John Benjamins B.V., p. 119-142.
- Berner, S., 2003, "Lost In Translation: Cross-Lingual Communication, And Virtual Academic Communities." <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.130.2973> (junio de 2010).
- Dale van, 2003, *Groot Woordenboeken Nederlands-Spaans & Spaans-Nederlands*, Utrecht, Van Dale Lexicografie.
- Chesterman, 2001, "Functional theories of translation". En: Munday, *Introducing Translation studies: theories and applications*, Routledge, Abingdon, p. 73-74.
- Chomsky, N., 1977, *Problemas actuales en teoría lingüística: temas teóricos de gramática generativa*. México DF, Siglo veintiuno.
- Colorado, N., Borja de, F., 2007, "Ingeniería del Lenguaje Natural", Universidad de Alicante. <http://rua.ua.es/dspace/handle/10045/3879> (junio de 2010).
- Cuetara Priede, J., 2007, "Manual de gramática histórica", Universidad Nacional Autónoma de México, <http://www.scribd.com/doc/26001579/Manual-de-Gramatica-Historica> (junio de 2010).
- Hockett, C., 1960, "The origin of speech", *Scientific American* 203, p. 88-96.
- Holmes, S., 2004, "De brug bij Bommel herbouwen". En: Naaijken, T., Koster, C., et al., *Denken over vertalen*, Nijmegen, Uitgever Vantilt, p. 273-287.
- Hutchins, J., 2003, "The history of machine translation in a nutshell" <http://www.hutchinsweb.me.uk/Nutshell-2005.pdf> (julio de 2010)
- Hutchins, J., 2007, "Machine translation: a concise history", Chinese University of Hong Kong <http://www.hutchinsweb.me.uk/CUHK-2006.pdf> (julio de 2010)
- Krauwier, S., 2003, "Aspecten van het automatisch vertalen", Utrechts instituut voor Linguïstiek Uil OTS. <http://www-sk.let.uu.nl/ond/mt.html> (julio de 2010)
- Krings, H., 2001, *Repairing texts: Emperical investigations of machine translation post-editing processes*, The Kent State University Press, Ohio.
- Linn, S., Slager, M., 2007, *Vertalen uit het Spaans: tekst en uitleg*. Bussum, Uitgeverij Coutinho.
- Munday, J., 2001, *Introducing Translation studies: theories and applications*, Routledge, Abingdon.
- Nirenburg, S. 1987, "Knowledge and choices in machine translation". En: *Machine translation: theoretical and methodological issues*, Cambridge, Cambridge University Press, p. 68-90.

- Reiss, K., 1976, *Texttyp und Übersetzungsmethode*, Kronberg, Scriptor Verlag GmbH & CO.
- Real Academia Española, 2001, *Diccionario de la Lengua Española*, Madrid, RAE, 22.^a ed.
- Somers, H., 2003, *Machine translation: latest developments*. En: Mitkov (ed), *Oxford handbook of computational linguistics*. Oxford, Oxford University Press, p.512-528.
- Vázquez Ayora, G., 1977, *Introducción a la traductología*, Georgetown, Georgetown University Press.

Páginas web

- <http://googleresearch.blogspot.com/2006/04/statistical-machine-translation-live.html> (junio de 2010)
- http://utrecht.cervantes.es/nl/algemene_cultuur_spaans/algemen_informatie_cultuur_spaans.htm (julio de 2010)
- <http://www.bcn.cat/climatechange/es/informacio-turistica.html> (julio de 2010)
- http://www.sgel.es/ele/ficheros/experiencias/Unidad_did%C3%A1ctica_40.pdf (julio de 2010)
- <http://www.carcajadasonline.com/erase-una-vez-un-borracho/> (julio de 2010)

Anexos
