



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

La posesición de traducciones de textos técnicos del alemán al castellano



**Universitat Autònoma
de Barcelona**

Tesis Doctoral

María Helena Mendoza García

Junio 2017

Dirigida por los Doctores:

Pilar Sánchez Gijón

Adrià Martín Mor

Doctorat de Traducció i Estudis Interculturals

Departament de Traducció i Interpretació i d'Estudios de l'Àsia Oriental

Universitat Autònoma de Barcelona

RESUMEN Y PALABRAS CLAVES

Resumen

Esta tesis pretende realizar un estudio descriptivo de la problemática de la traducción de textos técnicos del alemán al castellano mediante traducción automática y elaborar una guía de ayuda para la posesición de dichas traducciones.

La hipótesis parte del hecho de que el resultado de las traducciones automáticas hasta el momento no es satisfactorio; que presentan carencias que sólo se pueden solucionar mediante el proceso de la posesición. Nuestra hipótesis consiste en que es posible sistematizar el tipo de problemas de traducción provocados por el uso de la traducción automática en una combinación lingüística y un género determinados y que la identificación de estos problemas permitirá elaborar una guía de directrices para la posesición.

Palabras claves

Traductología, traducción automática, tecnologías de la traducción, posesición, errores en la traducción, problemas de la traducción, directrices para la posesición, motores de traducción automática, control de calidad de la traducción, formación del poseedor, ayuda en la posesición.

ABSTRACT AND KEY WORDS

Abstract

At present, automatic translations of technical texts from one language to another show deficiencies that can only be solved through a post-editing process.

This Thesis intends to study the problems encountered in the machine translation of technical texts from German into Spanish and create a practical guide that helps with the post-editing of those translations. We hypothesize that it is possible to implement a system to identify the problems found in an automatic translation in a linguistic combination and a settled genre that would allow us to create a coherent, established guide to help with the post-editing process.

Key words

Translation studies, translation technologies, machine translation, post-editing, translation errors, translation problems, guidelines for post-edition, automatic translation engines, translation quality control, post-editor training, post-edition guidelines

LISTA DE ABREVIATURAS

BLEU	BiLingual Evaluation Understudy
CAT	Computer-Aided Translation
DQF	Dynamic Quality Framework
DFKI	Deutsche Forschungszentrum für Künstliche Intelligenz
EBMT	Example Based Machine Translation
FAHQT	Fully Automatic High Quality Translation
HAMT	Human Aided Machine Translation
LISA	Localization Industry Standards Association
LV	Distancia de teclado Levenshtein
LSP	Language Service Provider
MAHT	Machine Aided Human Translation
MT	Machine Translation
MQM	Multidimensional Quality Metrics
NIST	National Institute of Standards
PE	Posedición
RBMT	Rule-Based Machine Translation
SGET	Sistemas de Gestión y de Edición de Traducciones
SMT	Statistical Machine Translation
TA	Traducción automática
TAO	Traducción Asistida por Ordenador
TAP	Think Aloud Protocol
TAUS	Translation Automaton User Society
TER	Translation Edit Rate
TM	Translation Memory
WER	Word Error Rate

ÍNDICE

Resumen y palabras claves.....	i
Abstract and key words	ii
Lista de abreviaturas.....	iii
Índice.....	v
INTRODUCCIÓN	1
PARTE I: MARCO TEÓRICO.....	7
1. Traducción, traducción técnica, problemas de la traducción.....	8
1.1 Definición de traducción	8
1.1.1 La equivalencia en la traducción	16
1.1.2 La finalidad de la traducción	22
1.2 La traducción técnica.....	24
1.2.1 Definición del texto técnico.....	26
1.2.2 Características específicas de la traducción técnica	28
1.3 Problemas de la traducción.....	30
1.3.1 Los problemas de la traducción según Nord	31
1.3.2 Los problemas de la traducción según Krings.....	35
1.3.3 Los problemas de la traducción según Schmitt	36
1.3.4 Los problemas de la traducción según Hurtado.....	38
1.4 Errores en la traducción	40
1.4.1 La clasificación de errores de traducción según Nord.....	41
1.4.2 La clasificación de los errores de traducción según Hurtado	42
1.4.3 La clasificación de los errores de traducción según Schmitt.....	43
1.5 Problemas de traducción en la combinación lingüística alemán-castellano.....	47
1.5.1 Problemas pragmáticos.....	48
1.5.2 Problemas culturales.....	49
1.5.3 Problemas en el uso y aplicación de convenciones	50
1.5.4 Problemas lingüísticos.....	53
1.6 Conclusiones del capítulo	72
2. Tecnologías de la traducción, traducción automática y sus problemas.....	73

Índice

2.1	Traducción asistida por ordenador	76
2.2	La traducción automática.....	79
2.2.1	Sistemas de traducción automática basados en reglas.....	81
2.2.2	Sistemas basados en datos o corpus	85
2.2.3	Sistemas de traducción automática basados en redes neuronales.....	98
2.2.4	Problemas y errores se presentan en la traducción automática	99
2.2.5	Métodos de evaluación de la calidad de la traducción automática.....	140
2.3	Conclusiones del capítulo	171
3.	Posedición.....	173
3.1	Definición	173
3.2	Tipos de posesición	175
3.3	Niveles de posesición	176
3.4	Reglas de posesición	178
3.4.1	Reglas para la posesición de TAUS	179
3.4.2	Directrices de posesición para GALE	181
3.4.3	Directrices dependiendo del tipo de posesición	182
3.5	Métodos de evaluación del trabajo de posesición	185
3.5.1	Medición del esfuerzo cognitivo	186
3.5.2	Medición del esfuerzo en parámetros de tiempo	187
3.5.3	Medición del esfuerzo técnico	187
3.6	Herramientas.....	188
3.7	Conclusiones del capítulo	189
4.	Conclusiones de la primera parte.....	190
PARTE II: MARCO METODOLÓGICO.....		201
5.	Presentación de los objetivos.....	202
5.1	Objetivos.....	202
5.2	Aplicación de sistemas de medición de la calidad como instrumento de investigación: el caso de MQM	204
5.2.1	Definición de calidad y errores en MQM.....	205
5.2.2	Tipos de errores	206
5.2.3	Scoring.....	207

Índice

5.2.4	Utilización de MQM.....	208
5.2.5	Cómo crear una métrica con MQM.....	208
6.	Descripción del corpus y de las herramientas de análisis.....	214
6.1	Descripción del corpus	214
6.1.1	Estructura del corpus extraído de TaraXÚ	217
6.1.2	Reorganización de los datos del corpus.....	220
6.2	Herramientas adicionales utilizadas	224
6.2.1	Árboles de decisiones	225
6.2.2	translate5.....	228
7.	Análisis preliminar del corpus y descripción de los motores de traducción	232
7.1	Categorización de los errores	234
7.2	Problemas y errores encontrados en los segmentos originales.....	236
7.3	Tipos de errores encontrados en las traducciones	238
7.3.1	Descripción global de los errores en las traducciones.....	238
7.3.2	Tipos de errores según categorías gramaticales	246
7.4	Errores encontrados según el motor empleado.....	255
7.4.1	Errores encontrados en las propuestas del motor SMT1	255
7.4.2	Errores encontrados en las propuestas del motor SMT2	257
7.4.3	Errores encontrados en las propuestas del motor RBMT1	259
7.4.4	Errores que se encuentran en las propuestas del motor RBMT2.....	261
7.5	Comparación de los sistemas.....	264
7.6	Análisis de las traducciones con ayuda de la distancia de teclado (Levenshtein) ...	265
7.7	Errores presentes en el corpus según las categorías de MQM.....	268
7.7.1	Errores encontrados en la categoría Accuracy	268
7.7.2	Errores encontrados en la categoría Design	272
7.7.3	Errores encontrados en la categoría Fluency	274
7.7.4	Errores encontrados en la categoría Internationalization	277
7.7.5	Errores encontrados en la categoría Locale convention	278
7.7.6	Errores encontrados en la categoría Style.....	278
7.7.7	Errores encontrados en la categoría Terminology	279
7.7.8	Errores encontrados en la categoría Verity.....	280
7.8	Conclusiones del primer análisis	281

Índice

8.	Clasificación de errores para la sistematización de la posesición.....	283
8.1	Procesamiento de los segmentos poseídos a partir de TA.....	283
8.2	Descripción de los elementos de la clasificación de errores.....	285
8.2.1	Macro categorías y categorías de errores.....	285
8.2.2	Valores de los errores.....	288
8.3	Definiciones por macro categorías y categorías de errores.....	288
8.3.1	Primera macro categoría: errores independientes de la combinación lingüística.....	289
8.3.2	Segunda macro categoría: errores relacionados con las características propias del alemán.....	292
8.3.3	Tercera macro categoría: errores relacionados con las características del castellano.....	296
8.3.4	Cuarta macro categoría: errores en la microestructura.....	301
8.3.5	Quinta macro categoría: errores relacionados con el género y funcionalidad del texto o la redacción del original.....	308
8.3.6	Sexta macro categoría: otro tipo de errores.....	310
8.4	Resultados globales según el tipo de motor.....	311
8.4.1	Errores encontrados según el motor empleado.....	318
8.4.2	Errores encontrados en las propuestas del motor SMT1.....	318
8.4.3	Errores encontrados en las propuestas del motor SMT2.....	320
8.4.4	Errores encontrados en las propuestas del motor RBMT1.....	323
8.4.5	Errores encontrados en las propuestas del motor RBMT2.....	326
8.5	Conclusiones del segundo análisis.....	333
PARTE III: RESULTADOS.....		337
9.	Incidencia de cada categoría de errores en el corpus de análisis.....	338
9.1	Incidenias por macro categorías.....	338
9.1.1	Incidencia de los errores en la primera macro categoría.....	340
9.1.2	Incidencia de los errores en la segunda macro categoría.....	342
9.1.3	Incidencia de los errores de la tercera macro categoría.....	343
9.1.4	Incidenias de los errores de la cuarta macro categoría.....	344
9.1.5	Incidencia de los errores de la quinta macro categoría.....	347
9.1.6	Incidencia de los errores de la sexta macro categoría.....	349

Índice

9.2	Incidencia por macrocategoría con relación al tipo de motor	350
9.2.1	Incidencia por tipos de motor en la primera macrocategoría	352
9.2.2	Incidencia por tipos de motor en la segunda macrocategoría.....	353
9.2.3	Incidencia por tipos de motor en la tercera macrocategoría	355
9.2.4	Incidencia por tipos de motor en la cuarta macrocategoría	356
9.2.5	Incidencia por tipos de motor en la quinta macrocategoría.....	358
9.3	Incidencia por categorías y motor concreto.....	359
9.3.1	Primera macrocategoría: errores independientes de la combinación lingüística.....	360
9.3.2	Segunda macrocategoría errores provocados por las características propias del alemán.....	366
9.3.3	Tercera macrocategoría: errores provocados por características del castellano	371
9.3.4	Cuarta macrocategoría: errores en la microestructura	376
9.3.5	Quinta macrocategoría errores relacionados con género del texto o la redacción del texto original	388
9.4	Incidencia por motor y valores	390
9.4.1	Valores asignados a las ediciones de verbos simples y compuestos	391
9.4.2	Valores asignados a las ediciones de sustantivos	392
9.4.3	Compuestos nominales	397
9.4.4	Verbos separables	402
9.5	Sugerencias	405
9.5.1	Sugerencias relativas a las fases anteriores a la posesición	406
9.5.2	Sugerencias a las fases a realizar durante la posesición.....	407
9.5.3	Sugerencias relativas a las fases posteriores a la posesición.....	407
PARTE IV: CONCLUSIONES		411
10.	Conclusiones.....	412
BIBLIOGRAFÍA.....		417
ÍNDICE DE ILUSTRACIONES Y TABLAS		429
PARTE V: ANEXOS		441

Índice

INTRODUCCIÓN

La investigación en el desarrollo de la traducción automática parece arrojar cada vez mejores resultados, lo que sin duda ha influido el hecho de que se ha incrementado su utilización en los últimos años. Los motores de traducción automática se continúan desarrollando con el fin de establecer cada vez una mayor concordancia entre el original y su traducción. Hasta el momento ninguno de los motores disponibles puede prescindir de la ayuda humana, ya sea en la fase de preparación de los textos originales y de los recursos necesarios en el proceso de traducción, en el entrenamiento de los motores, en la evaluación del rendimiento de los sistemas o en la posesición de sus resultados.

En esta tesis nos dedicaremos a este último aspecto, a la posesición como el paso o conjunto de pasos en un proceso de traducción ya iniciado, en el que un poseedor modifica, edita y/o corrige los textos que han sido traducidos con ayuda de un motor de traducción automática.

El tema central de nuestra tesis son las nociones de problemas y errores de traducción que se dan particularmente a partir del uso de traducción automática estadística y basada en reglas. Si bien la clasificación de los errores que se encuentran en la traducción automática puede tener como fin estimar la calidad del texto obtenido mediante traducción automática, nuestro interés está dirigido a analizar los errores presentados para facilitar —como veremos más adelante— el proceso de posesición y de tal manera a entender cómo se puede mejorar y qué tan exigente es cada tipo de solución. Nuestro interés no es de índole técnico, no pretendemos establecer directrices que mejoren el desarrollo de los sistemas o motores de traducción automática. Nuestro objetivo es establecer parámetros que puedan ayudar al poseedor y

Introducción

contribuir de esta manera en la formación académica de los traductores, que hoy en día no pueden dejar de lado el auge de los sistemas automáticos y deben ser preparados para poder ampliar su competencia como traductores.

Existen otros acercamientos a la evaluación de errores en la traducción automática. La evaluación manual de errores de la traducción automática conlleva más tiempo y esfuerzo ya que se debe instruir a los evaluadores, se requiere de directrices detalladas y concisas de evaluación y por último —pero no por ello menos importante— de programas que permitan a los evaluadores la anotación para hacer su trabajo de forma rápida y eficiente, ya que cuando hay muchos textos los evaluadores pueden perder el foco.

También nos encontramos con métricas basadas en juicios y en escalas más o menos subjetivas en las que se trata de medir grados de fluidez y de adecuación. Los resultados de las métricas que utilizan clasificaciones como la 'mejor traducción' son también relativas, pues no existe un criterio absoluto para definir qué es mejor o más adecuado, sea cual sea su contexto.

Objetivos

El principal objetivo de esta tesis consiste en proporcionar una visión pormenorizada de la posesición de textos técnicos del alemán al español, a partir del análisis de un corpus paralelo de segmentos alineados con sus traducciones automáticas en cuatro motores y sus posesiciones publicadas. El cumplimiento de este objetivo permitirá identificar criterios que favorezcan la creación de estrategias para la formación de poseedores teniendo en cuenta tanto la combinación lingüística como las circunstancias habituales de un encargo de posesición. De este modo, los resultados de esta tesis se podrán convertir en un punto de partida para la formación académica de poseedores.

Así cumplir con este objetivo debemos plantearnos una serie de objetivos previos o parciales, que serían:

1. Describir el corpus de segmentos originales y sus traducciones con cuatro motores diferentes que utilizaremos como punto de partida para este análisis y describiremos en el apartado “Descripción del corpus” en la página 214.

Introducción

2. Analizar los resultados de cada motor.
3. Establecer una categorización de errores de acuerdo con las ediciones introducidas durante la fase de posesición, así como con su segmento origen.
4. Analizar los resultados teniendo en cuenta las categorías de errores establecidas y el motor de traducción automática que produjo el segmento elegido por el poseedor.
5. Proponer sugerencias que puedan implementarse tanto en contextos de formación de poseedores como en proyectos de posesición profesionales.

Nuestro objetivo es abordar los textos y los errores de la traducción automática teniendo en cuenta también los textos de partida y viendo cómo se relacionan. En este sentido, nuestro enfoque difiere de otros sistemas de caracterización de errores o de análisis de la calidad, que tienen en cuenta únicamente el texto de llegada, porque nuestro interés no es mejorar los resultados de los sistemas de traducción automática sino hacer un aporte en la formación de traductores y poseedores.

Queremos ver si podemos realizar una clasificación de errores que ayuden al poseedor en nuestra combinación lingüística; pretendemos analizar si entre los errores y los motores empleados existe algún tipo de relación que pueda aportar algún tipo de ayuda en la formación de traductores o poseedores con la información obtenida. Como indicamos en la introducción de nuestra tesis el objetivo es profundizar en la descripción de los errores en que se dan en la traducción automática en textos de carácter técnico para facilitar el establecimiento de estrategias y técnicas de posesición.

Metodología

Para poder evaluar los textos del corpus necesitamos un listado de los errores que pueden presentarse en la traducción. Los errores en la traducción se producen cuando no se solucionan correctamente los problemas que se presentan en el proceso de traducción. También nos percatamos de la existencia de los problemas de la traducción cuando vemos los errores que se dan, es decir, cuando éstos no se solucionan.

A partir de la clasificación que daremos a continuación podremos establecer más adelante qué tipos de errores de traducción consideramos relevantes para la evaluación de la traducción de los textos que aparecen en el corpus.

Introducción

Los problemas nos dan una base teórica para analizar el corpus. Al clasificar los errores podemos ver dónde realmente se equivoca la máquina y es por ello que para la evaluación de la traducción automática la clasificación de errores tiene un papel tan importante.

Para cumplir el objetivo que nos hemos fijado haremos un estudio basado en corpus y en la propuesta de errores desarrollada en el marco teórico determinar qué tipos de errores se presentan en un caso real y si el análisis de estos datos nos permite desarrollar directrices que ayuden al poseedor. Trataremos de crear un vínculo entre las categorías de errores que establecimos a partir del marco teórico, es decir basadas en los postulados de los diferentes investigadores, y las evaluaciones prácticas de las traducciones automáticas en términos de precisión o coincidencia con el original y fluidez en el idioma meta.

Estructura

Para abordar la temática hemos estructurado la tesis en cuatro partes principales: el marco teórico (capítulos 2 a 5), el marco metodológico (capítulos 5-6), en tercera parte daremos una descripción detallada de los resultados (capítulos 7-9) y finalmente en la última parte (capítulo 10) presentaremos nuestras sugerencias.

En el marco teórico estableceremos, por medio de una revisión de la bibliografía, las bases teóricas de los conceptos manejados en la tesis. En primer lugar en el capítulo 1, trataremos las definiciones de traducción, traducción técnica, errores y problemas de la traducción así como los problemas que suelen presentarse en nuestra combinación lingüística. En este capítulo también se tratarán cuestiones como la definición de equivalencia y finalidad en la traducción, la definición del tipo de texto técnico y las características específicas de la traducción técnica. En el capítulo siguiente (2), trataremos el otro aspecto principal de la tesis: las tecnologías de la traducción. Este capítulo se divide principalmente en dos bloques principales: la traducción asistida por ordenador y la traducción automática. El primer bloque tratará sobre todo las memorias de traducción desde un punto de vista general, mientras que el segundo bloque se describirá más detalladamente el tema de la traducción automática y abordaremos también los tipos de problemas que —desde el punto de vista teórico— se pueden presentar cuando se emplean sistemas de traducción automática. Así, obtendremos la información necesaria para establecer los aspectos a partir de los cuales se deberá observar y com-

Introducción

rar los datos del corpus. En el capítulo 3 introduciremos el concepto de posesición, sus diferentes tipos y los métodos de evaluación de la calidad tanto del trabajo mismo de posesición, como de la calidad de la traducción automática. Al final del marco teórico habremos descrito los conceptos fundamentales que nos ayudarán en el análisis de los datos del corpus y obtendremos los rasgos más relevantes para analizar y medir estos datos.

Después de haber aclarado desde el punto de vista teórico la definición general de los problemas y errores y de cuáles de estos son los puntos más importantes a tener en cuenta en la combinación lingüística alemán-castellano, debemos preguntarnos ahora si la definición teórica dada se ajusta al contexto particular de la traducción automática en nuestra combinación lingüística. Para poder abordar el problema central de esta tesis, es decir, el análisis de los problemas que se pueden presentar en la traducción automática, necesitaremos una clasificación general, que no se guíe en las habilidades y competencias del traductor o poseedor, sino en las características mismas del proceso de traducción y de los idiomas involucrados en este proceso. Para desarrollar este objetivo hemos realizado en la segunda parte de la tesis, luego de la presentación de nuestros objetivos (capítulo 5), una descripción del corpus y de las herramientas utilizadas (capítulo 6). El capítulo 7 tendrá como tema central el análisis preliminar de los datos del corpus y la descripción general de los problemas encontrados en todas las propuestas de traducción, incluso aquellas que no se poseitaron. En este capítulo analizaremos los datos del corpus también teniendo en cuenta también la clasificación de errores de la métrica MQM. Con base en el análisis preliminar y el análisis según la tipología de errores de MQM presentaremos en capítulo 8 nuestra clasificación de errores y los resultados globales de aplicar esta clasificación en teniendo en cuenta las propuestas seleccionadas para la posesición. La clasificación de errores propuesta aquí tiene como fin contribuir en la formación académica del traductor como poseedor.

En la tercera parte evaluaremos los resultados del análisis del corpus teniendo según nuestra definición en macrocategorías de errores y valores de errores; con ayuda de esta evaluación haremos finalmente nuestras sugerencias.

Introducción

PARTE I: MARCO TEÓRICO

El marco teórico está dividido en cuatro capítulos en los que, por medio de un análisis de la bibliografía, nos referiremos a los conceptos teóricos fundamentales que trataremos en nuestra tesis, es decir a las definiciones actuales de los conceptos de traducción, traducción técnica, problemas y errores en la traducción. A continuación, en el capítulo 3 nos referiremos a las tecnologías de la traducción, a la traducción automática y a los problemas y errores que se dan en la traducción automática. En el capítulo 4 introduciremos el concepto de posesión, sus diferentes tipos, reglas, métodos de evaluación y herramientas, también por medio de una revisión de la bibliografía. Finalmente en el capítulo 5, presentamos un resumen de los problemas y errores propuestos encontrados en el análisis bibliográfico, que nos ha de servir para el desarrollo de la tesis y de la segunda parte de esta.

1. Traducción, traducción técnica, problemas de la traducción

Es este capítulo expondremos en primer lugar las definiciones de tres conceptos fundamentales en nuestra tesis. En primer lugar nos referiremos a qué entendemos por traducción, sin pretender un análisis exhaustivo expondremos algunas de las definiciones y reflexiones contemporáneas sobre la traducción y el proceso de traducir. Ya que en nuestra tesis no nos ocuparemos de la traducción en general, sino de la traducción técnica especificaremos en qué sentido hablaremos de este tipo de traducción y por último mencionaremos cuatro clasificaciones de problemas de la traducción que son relevantes para el tratamiento de la traducción técnica en nuestra tesis

1.1 Definición de traducción

Formular una correcta definición de un término simple no es una tarea fácil, a menudo suelen presentarse dificultades que pueden ser de un grado mucho mayor si el término a definir representa un fenómeno complejo. Para hacer frente a estas dificultades nos encontramos con reglas clásicas que fueron postuladas para facilitar la tarea de definir y que recomendaban el establecer el *genus proximum et differentia specifica*. En las diferentes definiciones contemporáneas del término traducción encontramos repetidamente ciertos *termini* como *transformación*, *sustitución*, *ilustración* e incluso *imitación* que podríamos incluir en la concepción clásica de *genus proximum*. En cuanto a la *differentia specifica* en las definiciones de traducción no encontramos un único rasgo común; solo nos encontramos con la caracterización de que en la traducción deben tomar parte mínimo dos idiomas y dos culturas. Ahora ¿qué elementos invariantes deben darse para que podamos hablar de traducción? Aquí nos encontramos con *termini* como significado, estilo, función, valor comunicativo e incluso efecto o sentido, términos que dependiendo del autor tienen un rango o prioridad diferente en la definición (Albrecht 2005, 25–27).

Algunas teorías no se ocupan directamente de la definición sino que se refieren a los modelos de traducción y hacen una diferenciación entre los procesos de traducción. En este sentido podemos hablar de dos tipos diferentes de definiciones:

1. Traducción, traducción técnica, problemas de la traducción

- a) Aquellas que ven el proceso de traducción de manera abstracta y estática; en estos modelos los factores implicados y las relaciones entre ellos se representan sin que se considere un proceso de traducción real.
- b) También nos encontramos con los modelos que ven el proceso de traducción como un proceso concreto, dinámico y psicológico y que pretende aclarar qué pasa realmente en el proceso o tarea concreta de traducir (Albrecht 2005, 27).

A pesar de las diferentes definiciones de lo qué significa traducir y de lo qué se puede entender como traducción nos encontramos con la premisa generalmente aceptada de que la traducción es una actividad o proceso necesario, porque existen diferencias entre lenguas y culturas que solo a través de la traducción pueden ser superadas. La finalidad de la traducción es —en este sentido— establecer la comunicación, que no sería posible de otra manera, debido a dichas diferencias. Es decir, la traducción está dirigida a un destinatario que desconoce la lengua y/o la cultura en que se produjo el texto o discurso original (Hurtado 2011, 28–29).

Hurtado define traducción como una habilidad, un *saber hacer*, en otras palabras un conocimiento esencialmente de tipo operativo que se adquiere fundamentalmente por la práctica (2011, 25). Otros autores como Koller (2011, 76) definen la traducción como un proceso de transformación de un texto escrito en un idioma a un texto en otro idioma. Para autores como Nord la traducción es una interacción comunicativa intercultural mediada y basada en un texto (Nord 2009, 212).

Estas y todas las otras definiciones contemporáneas de traducir y traducción parten de una distinción previa. Hablamos aquí de la distinción entre los tipos de traducciones que define por primera vez Jakobson (1959, 233). Para este autor podemos hablar de tres tipos fundamentales de traducción: la traducción intersemiótica, que es la interpretación de signos verbales mediante signos no verbales, por ejemplo si se *traduce* una guía de instrucciones escrita en texto a una guía que dé las instrucciones con ayuda de dibujos y gráficas. La traducción intralingual o formulación, que es una interpretación de ciertos signos verbales mediante otros de la misma lengua; este tipo de traducción se da si se traduce un texto por ejemplo escrito para el personal médico a un lenguaje que pueda entender un paciente, es decir una persona que no tenga conocimientos especializados en el tema. Por último la traducción interlingüísti-

1. Traducción, traducción técnica, problemas de la traducción

ca, que se define como la interpretación de los signos verbales de una lengua en otra, es decir, la definición que Jakobson denomina como *translation proper* y el tipo de traducción al que nos referiremos en nuestra tesis.

Si nos referiremos a la traducción como la interpretación de los signos verbales de una lengua en otra, nos encontramos también con diferentes teorías y definiciones de lo que se pueden entender como traducción. Hurtado clasifica cuatro tipos principales de definiciones de la traducción:

1) Aquellas definiciones como la postulada por Vinay y Darbelnet quienes consideran la traducción como una actividad entre lenguas, es decir como el “pasar de una lengua A a una lengua B para expresar la misma realidad”. Esta definición solo tiene en cuenta los elementos lingüísticos, sitúa la traducción en el plano de la lengua y no del habla y no da una definición clara de lo que se entiende como *realidad*, a pesar de ser esto mismo lo que permanece en el proceso de traducción (Vinay y Darbelnet 1958).

2) En el otro grupo que menciona Hurtado nos encontramos con definiciones de autores como Seleskovitch y Lederer que entienden la traducción como una actividad textual. Para estas teorías traducir significa transmitir el sentido de los mensajes que contiene un texto y no convertir en otra lengua la lengua en la que éste está formulado (2011, 38). Este tipo de definiciones no solo hacen énfasis en el carácter textual de la traducción y sino que acentúan además el hecho de que el objeto de la traducción es el transmitir el sentido. Entre los autores que resaltan el aspecto textual de la traducción encontramos también a Catford (1965) y a House (1977, 29).

3) Existen teorías que resaltan el papel de la traducción como un acto comunicativo, en el que influye el contexto sociocultural. Este tipo de perspectiva la comparten teóricos como Nida y Taber (1969, 29) quienes introducen en su definición el concepto de equivalencia afirmando que la traducción consiste “... *in reproducing in the receptor language the closest natural equivalent of source – language message, first in terms of meaning and secondly in terms of style*” (1969, 12). Hatim y Mason plantean la traducción como “un proceso comunicativo que tiene lugar en un contexto social” (1990, 13) y Hermans (1991) concibe la traducción como una práctica comunicativa y, por lo tanto, un tipo de comportamiento social.

1. Traducción, traducción técnica, problemas de la traducción

Siguiendo esta orientación definen Vannerem y Snell Hornby (1986, 189–199) la traducción como un proceso en dos fases. Al comprender el texto a traducir el traductor parte de un marco dado: el texto y sus componentes lingüísticos. Dicho texto fue creado por un autor desde su propia experiencia, antecedentes, con un repertorio típico de escenas prototípicas. Por otro lado el marco global del texto (y todas las *frames* principales y secundarias dentro del texto) generan una reacción cognitiva en la imaginación del lector. Partiendo de estos escenarios o situaciones el traductor debe buscar en la lengua meta los escenarios o *frames* que creen en el receptor de la traducción la misma reacción.

Reiss y Vermeer realzan la finalidad como el principio dominante de toda traducción (1996, 82) y Nord profundiza esta tradición afirmando además que la traducción es un acto comunicativo cuyos criterios primordiales son la funcionalidad y la lealtad (2009, 219–220).

4) En el cuarto grupo de definiciones que menciona Hurtado encontramos aquellas definiciones en las que la traducción es vista como proceso. La traducción es el proceso de analizar el texto original, trasladar las oraciones de la lengua original en oraciones equivalentes en la lengua término para después ajustarlas estilísticamente (Vásquez-Ayora 1977, 50).

En esta línea hallamos autores como Seleskovitch y Lederer quienes ven la traducción como un proceso no de comparación entre dos lenguas, sino en el que se relacionan los procesos de comprensión y reexpresión (1984, 18).

Por otra parte Koller (2011) hace una clasificación de definiciones de la traducción en la que reconoce que a pesar de que existen muchas dilucidaciones en las que se tematizan las diversas condiciones y factores tanto internas como externas de la traducción, dichas definiciones se pueden agrupar en dos grupos principales: aquellas que definen la traducción desde la perspectiva textual y aquellas que la definen como un proceso.

En el primer grupo de definiciones que especifica Koller, es decir en el grupo de las teorías en las que el texto y el lenguaje ocupan el papel principal, nos encontramos las definiciones de autores como Oettinger (1960, 110). Para Oettinger la traducción es el proceso de transformación y sustitución de signos y representaciones en un lenguaje, por signos y representaciones equivalentes en otro, dónde lo más importante es la preservación del sentido ori-

1. Traducción, traducción técnica, problemas de la traducción

ginal. El concepto de equivalencia de sentido ocupa aquí un lugar importante en el proceso de traducción. En este grupo sitúa Koller también la definición de Catford (1965), y en esto coincide con Hurtado, ya que para el autor no es el sentido, sino el texto mismo el que ocupa el papel principal en el proceso de traducción. Para Catford la traducción es una sustitución del texto del idioma base por un texto equivalente en el idioma meta. La equivalencia aquí se da a nivel del texto en los dos idiomas implicados y por ello es para este autor la equivalencia el criterio más importante en el proceso de traducción.

Winter (1961), quien también hace parte de esta corriente, incluye en su definición no solo el texto sino también el *universo*, o por así decirlo la situación cultural en la que se origina el texto mismo (1961, 68). En su definición del proceso de traducción nos encontramos con tres elementos centrales: el texto o formulación, la parte del universo y la comprensión de los elementos *intra* y *meta* lingüísticos.

El otro grupo de definiciones que distingue Koller es el de las teorías en las que el aspecto comunicativo de la traducción es interpretado como la tarea fundamental. En este grupo de teorías de la traducción encontramos la definición de Wills (1977, 72). Para Wills la traducción es un proceso que consta de dos fases principales: la fase de la interpretación y la fase de la reproducción. En el proceso de traducción el traductor tiene en primer lugar la tarea de analizar y comprender el sentido y la intención estilística del texto origen para así poder en segundo lugar reconstruirlo en la segunda fase y de esta manera poder producir en el texto origen un texto equivalente en sentido comunicativo.

Del mismo modo Jäger describe la traducción como la producción de un texto en el lenguaje meta que sea equivalente comunicativamente con el texto original. La esencia de la traducción es asegurar la comunicación, de tal manera que el valor comunicativo de un texto se mantenga y el texto original y su traducción puedan considerarse como equivalentes a nivel comunicativo. La esencia de la traducción, así como la de la comunicación en general se encuentra a nivel extralingüístico, aunque en el área lingüística (lengua) se lleve a cabo la traducción (Jäger 1975, 36).

Stolze (2011) hace otra clasificación que parte del punto de vista que se toma para abordar el tema de la traducción. Según este autor se puede hablar de teorías que parten del

1. Traducción, traducción técnica, problemas de la traducción

punto de vista de los sistemas del lenguaje, aquellas que se centran en el aspecto textual de la traducción, las que definen la traducción desde la perspectiva de la disciplina misma, las que la ven como una acción y aquellas que tratan de definir la traducción desde la perspectiva del traductor.

También Pym nos da una clasificación de las diferentes teorías de la traducción occidentales a partir de la segunda mitad del siglo XX, que según él se pueden ordenar en torno a cinco conceptos o paradigmas fundamentales explicativos: equivalencia, finalidad o *skopos*, descripción, indeterminismo y localización (Pym 2012, 7). Sea cual fuese el tipo de definición que se favorezca el hecho mismo de definir qué se entiende como traducción es definitivo también para poder determinar cómo medir la calidad de la traducción, en otras palabras, el tipo de definición que se escoja determinará los parámetros que definirán la calidad de la traducción (Melby et al. 2014). Melby y los coautores afirman que hay dos ejes principales con ayuda de los cuales se pueden describir las diversas definiciones de la traducción en la actualidad. Por un lado tenemos el tipo de alcance que puede tener la definición (estricta o amplia) y por otro las especificaciones de la traducción (absolutas o relativas). Según los autores se puede hablar de:

1. una definición de amplio alcance o *broad definition*, en la que las especificaciones puedan variar en relación con la audiencia y el propósito; y
2. una definición con un alcance limitado o *narrow definition* con que tiene un conjunto básico de especificaciones absolutas.

Desde esta perspectiva una definición amplia considera la traducción como “*the creation of target content that corresponds to source content according to agreed upon specifications*” (2014, 397). Este tipo de definición comenzó a desarrollarse a partir de la década de los 80 cuando la vinculación al proceso de la traducción de ciertas actividades, como la localización de software, involucraron otros elementos a tener en cuenta en la definición de la traducción, que iban más allá de la concepción de la traducción como una actividad centrada en el texto. Este tipo de elementos —por ejemplo la cultura, los usos regionales específicos, la música, elementos visuales, funcionales y no textuales— pueden variar e influir en la traducción. Es por ello que se pueden incluir en el concepto de traducción no solo los procesos y produc-

1. Traducción, traducción técnica, problemas de la traducción

tos que tienen que ver con la traducción humana profesional, sino también la traducción automática con o sin posesión, la localización, la traducción resumen, *gisting*, traducción técnica/científica, traducción legal, traducción literaria; transcreación; traducción abierta (extranjera, exotizada) o la traducción encubierta (domesticada).

Esta amplia definición de la traducción permite la traducción textual y adaptación no textual, siempre y cuando el contenido a traducir incluya algún texto. En esta definición — que se basa en el funcionalismo (Nord 1996)— la traducción es considerada globalmente y se puede llamar traducción *multimodal* para distinguirla de la traducción tradicional que se centra solo en el texto.

Debido a que en esta definición amplia más procesos y productos se califican como traducción el término ‘texto original’ no es suficiente para describir el antecedente de una traducción. Los términos ‘contenido original’ y ‘contenido meta’ son mucho más adecuados. De hecho, la industria de la lengua utiliza estos términos con frecuencia cada vez mayor.

A diferencia de la definición amplia, una definición estricta de la traducción hace una distinción entre *traducción per se* y otro tipo de actividades o *translation plus*. Bajo este punto de vista la traducción textual es la actividad central de los traductores profesionales, mientras que las actividades tales como traducción-resumen, localización, transcreación, y la traducción automática con posesión son habilidades complementarias que solo algunos traductores requieren (*translation plus*). Una definición estricta de la traducción afirma que:

Translation transfers a written source text into a written target text of roughly equivalent length. Such a translation conveys all the source text's meaning, making only those adjustments necessary for cultural appropriateness without adding, omitting, condensing, or adapting anything else (Melby et al. 2014, 397).

Como enunciamos anteriormente existe un segundo eje que es relevante para la definición de la traducción y es la medida o papel que se da a las especificaciones en el proceso de traducción. Las especificaciones definen lo que se espera en una traducción. Sin especificaciones los detalles de la traducción no se pueden determinar y puede darse lugar a confusión. Las partes interesadas emplean las especificaciones para dar forma a los contornos de una traducción en particular. El asunto principal aquí es si existen especificaciones absolutas para todas las traducciones o si las especificaciones deben variar en las tareas particulares.

1. Traducción, traducción técnica, problemas de la traducción

Dependiendo de si se opta por una definición amplia o una definición estricta de traducción se cambia también el valor que se otorga a las especificaciones, aunque ninguna de las posiciones duda de la importancia misma de estas, ya que sin especificaciones los traductores no pueden proporcionar la mejor traducción.

Estrechamente ligada a la definición amplia de la traducción es la noción de especificaciones de la traducción estructuradas en parámetros, que pueden variar independientemente (Melby et al. 2014, 402–403). Por ejemplo parámetros como idioma meta, público, propósito, registro, formato, diseño o maquetación, estilo o correspondencia con el contenido pueden tener valores diferentes dependiendo del proyecto.

En las especificaciones la importancia de los parámetros, por ejemplo estilo y correspondencia con el contenido, pueden variar considerablemente de un proyecto a otro. Lo que no varía es el papel que la precisión y la fluidez siguen teniendo para los parámetros en cuestión. En lo que se refiere al estilo, las especificaciones suponen que el aspecto textual del contenido meta debe tener un nivel mínimo de fluidez determinado por la audiencia y el propósito mismo del encargo. En cuanto a la correspondencia de contenido, las especificaciones suponen que el contenido debe tener un nivel mínimo de precisión, es decir debe existir una correspondencia entre el texto original y el texto meta, cuyo nivel también es determinado por la audiencia y el propósito del encargo de traducción.

Una definición amplia de traducción permitiría una variación considerable en la correspondencia de contenido en ciertos tipos de contratos de traducción, como por ejemplo el subtítulo de películas. Al igual que la precisión, el requisito de fluidez está ligado al eje del alcance. Una visión estricta de la traducción permite un requisito absoluto de la fluidez máxima para el texto de destino. Propuestas de traducción que no cumplan con este requisito absoluto serían consideradas como inaceptables. Por el contrario, una visión amplia de la traducción acepta —por ejemplo— una traducción automática en bruto, siempre y cuando el contenido cumpla con los propósitos de los usuarios.

La cuestión principal en relación con el eje de especificaciones es establecer si existen requisitos que se pueden considerar con antelación, ya que son requisitos válidos para todas las traducciones, independientemente de si se tiene una posición amplia o estricta en cuanto al

1. Traducción, traducción técnica, problemas de la traducción

alcance de la definición. Por lo tanto, la pregunta principal con respecto a las especificaciones es si existen absolutos a este nivel. En otras palabras, ¿existen requisitos que se puedan aplicar a todas las actividades de traducción? Los únicos requisitos aceptados de manera general son fluidez y precisión. En resumen en esta perspectiva la precisión está vinculada directamente al alcance de la definición: una visión estricta de la traducción exige la precisión absoluta, por el contrario en una definición amplia no se puede fijar un único estándar de precisión.

Como ya lo menciona Koller sea cual fuera la definición por la que nos inclinemos todas las teorías y sus clasificaciones tienen algo en común, al tratar de definir qué se entiende por traducción tienen un aspecto descriptivo, pero también tienen un aspecto normativo porque no solo describen el proceso sino que además tratan de formular cómo debe ser. Dicho aspecto se expresa en el concepto de equivalencia o en la pretensión de la equivalencia que tiene la traducción (Koller 2011, 91). La pretensión de equivalencia se puede referir a muchos parámetros diferentes: contenido, texto, estilo, normas de la lengua meta, función comunicativa en la lengua de origen, receptor, los escenarios o *frames* del texto original, etc.

1.1.1 La equivalencia en la traducción

No solo Koller menciona la importancia del concepto de equivalencia para la definición de traducción, como vimos en el capítulo anterior en varias de las definiciones de traducción se menciona repetidamente el concepto de equivalencia. La equivalencia presupone que un texto origen y un texto meta pueden tener el mismo valor, por lo menos en cierto nivel y en ciertos fragmentos. Por ello una traducción se diferencia de otro tipo de textos, porque tiene un valor compartido con el texto origen. El apogeo del concepto de equivalencia como ligado a la traducción comenzó con la lingüística estructuralista. Los primeros en utilizar el concepto de equivalencia para explicar la traducción son Vinay y Darbelnet (1958), Jakobson (1959) y Nida (1959). Mientras que para los primeros autores el concepto de equivalencia es solo un procedimiento de traducción más entre otros, para Nida y Jakobson es un principio básico y un problema fundamental; y como ya mencionamos, otros autores como Catford o Nida y Taber utilizan el concepto de equivalencia incluso para definir la traducción.

1. Traducción, traducción técnica, problemas de la traducción

No obstante, y a pesar de su importancia, no hay un consenso en la definición del término *equivalencia*. Nord afirma que equivalencia es uno de los conceptos más ambiguos y más diversamente interpretado en los estudios sobre traducción (Nord 1991, 25). Es debido precisamente a esta ambigüedad que a partir de los ochenta muchos autores como Snell-Hornby cuestionan la vigencia y utilidad del término *equivalencia*, e incluso habla de la ilusión de la equivalencia y considera el término como inapropiado para la teoría de la traducción, ya que no solo es impreciso y está mal definido sino que además "...presenta una ilusión de simetría entre lenguas que apenas existe más allá de un nivel de vagas aproximaciones y que tergiversa el problema básico de la traducción" (1988, 22). También otros teóricos como Gentzler (1993) consideran el concepto de equivalencia como perjudicial para la traductología.

Ante estas críticas y las preguntas acerca de la validez del término *equivalencia* para la traducción sugieren Reiss y Vermeer no renunciar al término, "sino precisar su contenido y restringir de manera adecuada su uso" (1996, 111). Es así como en este sentido afirma Hurtado (2011, 223) que hay que partir de una caracterización flexible y dinámica de la equivalencia considerándola como un concepto que define la relación que existe entre la traducción y el texto original. Dicha relación se da en función de una situación comunicativa, es decir tiene en cuenta el receptor y la finalidad de la traducción y está inmersa en el contexto socio-histórico en que se da el acto traductor y por ello mismo se caracteriza por ser una noción relativa, dinámica y funcional.

1.1.1.1 *Carácter dinámico de la equivalencia*

Existen unidades léxicas, frases, elementos morfosintácticos, gestos e incluso elementos culturales que pueden ser —a pesar de sus diferencias— directamente utilizados como equivalentes en las traducciones. Por ejemplo, mientras que en las culturas anglosajonas el *Friday the 13th* tiene la connotación de ser un día de mala suerte, en castellano tendría el martes 13 una significación equivalente (Pym 2012, 22). Esto se debe a que existen diversos factores que intervienen en la construcción de la equivalencia traductora y que le otorgan un carácter relativo:

1. Traducción, traducción técnica, problemas de la traducción

- El **contexto lingüístico y textual** en el que se insertan esos elementos hace que la traducción de una palabra o expresión pueda variar, dependiendo del contexto en que se encuentra. En castellano utilizamos la expresión “Te lo he dicho *mil* veces” como recriminación, mientras que en francés se utiliza *Je te l’ai dit cent fois*. Pym (2012, 23) también nos da un ejemplo de este tipo, en la televisión de los países de habla inglesa hay un conocido programa de concurso llamado *The Price is Right*. La adaptación de este programa en español se tituló *El precio es justo* y en francés *Le juste prix*, mientras que en alemán se tradujo como *Der Preis ist heiss* (El precio está caliente) ya que en la contexto lingüístico alemán la alusión a un juego infantil tenía más sentido en este idioma que la traducción *Der Preis ist gerecht*.
- El **tipo y género textual** en el que se encuentran. Aquí se tendrían en cuenta por ejemplo las convenciones del género en cuestión en ambas lenguas.
- El **contexto socio-histórico** en que se ubica la traducción.
- El **tipo de traducción** que se efectúa, es decir si es escrita, oral o audio-visual. No es lo mismo traducir por escrito una frase, donde el traductor tiene la posibilidad de hacer una descripción o una adaptación, que si se está haciendo una interpretación o una traducción para el doblaje de una película.

Estos factores hacen que la relación con el texto original sea diferente, que las equivalencias varíen y que el traductor utilice diversas técnicas según el caso, llegando a soluciones que tal vez solo puedan ser válidas para esa ocasión (Hurtado 2011, 211).

1.1.1.2 Clasificaciones de los tipos de equivalencia

Así como existen diferentes concepciones del concepto de equivalencia también existen varias clasificaciones de la equivalencia en traducción, que se distinguen según el punto de vista que toman:

- Las clasificaciones que se centran en el nivel en que debe darse la equivalencia, es decir, a nivel de las palabras, de la oración o del texto.
- Las clasificaciones que se basan en el tipo de significado reproducido, hablamos aquí de equivalencia denotativa, connotativa, pragmática, etc.

1. Traducción, traducción técnica, problemas de la traducción

- Las clasificaciones que parten del grado de equivalencia que se puede alcanzar, es decir si la equivalencia es aproximada o si no hay equivalencia, etc.

Muchas de estas clasificaciones se basan en criterios puramente lingüísticos y por ello se trata de clasificaciones de equivalencias entre lenguas, más no necesariamente de equivalencia traductora. Catford (1965) quien parte del concepto de equivalencia textual contraponiéndolo al de correspondencia formal, establece también una distinción entre equivalencia *nil*, que se da cuando no existen en la lengua meta ningún equivalente al de la lengua original y equivalencia *cero* que se da cuando el equivalente si existe, pero no es utilizado. Kade (1968, 79–83), propone una clasificación de equivalencia a nivel de palabras o frase de cuatro tipos:

- Equivalencia total o ideal, en la que se da una correspondencia ‘1:1’ tanto al nivel de la expresión como del contenido. Este es el nivel que encontramos en la traducción de nombres, títulos, denominaciones geográficas, términos técnicos estandarizados internacionalmente y números.
- Equivalencia facultativa, que se da cuando existen muchos equivalentes en la lengua meta y solo el contexto permite establecer el indicado. Aquí tenemos una equivalencia de ‘1:a varias posibilidades’ que es lo que sucede cuando los traductores tienen que elegir una solución entre varias alternativas.
- Equivalencia aproximada, que se da cuando la correspondencia semántica es parcial, es decir, se da una equivalencia ‘1:1’ a nivel de la expresión, pero de ‘1: correlación parcial’ a nivel del contenido.
- Equivalencia cero, se da cuando no hay un equivalente para una unidad léxica, es decir ‘1:0’ tanto a nivel de la expresión como del contenido.

Para Koller (2011) solo se puede hablar de equivalencia en una correspondencia 1:1 y distingue cinco tipos de equivalencia:

- Equivalencia denotativa, existe cuando el texto meta expresa los mismos hechos extralingüísticos que el texto original.

1. Traducción, traducción técnica, problemas de la traducción

- Equivalencia connotativa, que busca plasmar las connotaciones del texto referente al plano del estilo, sociolectos y diferencias geográficas y se da cuando la verbalización de elementos en el origen y el texto meta evocan reacciones emocionales y asociativas similares.
- Equivalencia normativa que persigue mantener la normativa lingüística y textual que rige determinados tipos de textos.
- Equivalencia pragmática o comunicativa, que procura que el texto meta produzca en el receptor de la traducción el mismo efecto que el texto original produjo en sus lectores.
- Equivalencia formal, que incluye las características formales, estéticas e individuales de los textos originales.

A estos tipos de equivalencias propuestos por Koller añade Königs (1981) otros dos que denomina equivalencias básicas:

- Equivalencia funcional o pretendida (*die textintendierte Äquivalenz*), que trata de mantener en la traducción la función del texto original.
- Equivalencia final, que atiende a la función del texto meta aunque a veces no coincidente con la función del texto original.

Hurtado afirma que dado el carácter contextual y dinámico de la equivalencia, cualquier tipo de clasificación es relativa. La textualidad y contextualidad de la equivalencia traductora anula aquellas clasificaciones que se basen en niveles (léxico, gramático, etc.) o que la desmenuzan en equivalencias parciales, es decir, equivalencia denotativa, connotativa, etc. La autora establece también que el concepto de equivalencia ha ido evolucionando en la traductología, desde una posición inicial que la situaba en el plano de la lengua, pasando por concepciones que la ven en el plano textual y contextual del habla hasta llegar a una concepción de equivalencia funcional y comunicativa (2011, 223).

Equivalencia en el plano de la lengua

En sus inicios el concepto de equivalencia tenía un carácter puramente lingüístico, y no se tenía en cuenta el plano contextual (Jakobson 1959, 70). Para autores como Vinay y Darbelnet (1958) la equivalencia es un procedimiento de traducción, tal como lo son el calco,

1. Traducción, traducción técnica, problemas de la traducción

la transposición o la adaptación. La equivalencia como procedimiento parte de la situación de comunicación y utiliza métodos estilísticos y estructurales individuales para cada lengua. Las equivalencias son fijas y pueden ser utilizadas repetidamente, por ejemplo en forma de proverbios, frases idiomáticas etc.

Catford (1965) hace una distinción entre correspondencia formal y equivalencia textual. Para Catford, un equivalente textual es cualquier forma del texto de la lengua de llegada que resulte ser el equivalente de una forma dada a la lengua origen (1970, 50). La correspondencia formal la define como “cualquier categoría de la lengua meta (unidad, clase, estructura, elemento de estructura, etc.) de la cual se puede decir que ocupa, tan aproximadamente como es posible, el ‘mismo’ lugar en la ‘economía’ de la lengua meta que el ocupado por la categoría en la lengua origen en la economía de la lengua origen” (1970, 49). Según la teoría de Catford el significado de un segmento en la lengua origen y el significado de un segmento en lengua meta no siempre coinciden. Sin embargo, aunque no coincidan formalmente ambos significados, sí pueden tener la misma función comunicativa. Así se puede afirmar que los segmentos son equivalentes a nivel textual y resaltar la importancia del contexto y de la situación comunicativa en cada lengua. Hurtado afirma que a pesar de que Catford hace consideraciones de índole textual los ejemplos que pone no son reales y no exceden el ámbito de la frase y por ello no trascienden el carácter textual y se quedan en el lingüístico (Hurtado 2011, 215).

Equivalencia en el plano del habla

Autores como Nida y Taber (1969) hacen una distinción entre equivalencia en el plano de la lengua y del habla y consideran esta última como la verdadera equivalencia de traducción. Con el concepto de *equivalencia dinámica* describen que la relación entre el receptor de la traducción y el mensaje traducido ha de ser la misma que la que existía entre el receptor y el mensaje original (Nida 1964, 159). El traductor debe buscar la equivalencia y no la identidad, es decir, debe reproducir el mensaje en vez de conservar la forma de las expresiones. Si el traductor se decide por la *correspondencia formal* se dificulta la comprensión del lector en la lengua meta.

1. Traducción, traducción técnica, problemas de la traducción

A partir de los años setenta se produce un cambio de paradigma teórico en la traductología que se aleja de la concepción lingüística del concepto de la equivalencia situado en la lengua por un concepto de equivalencia de carácter textual. Las equivalencias en la traducción son discursivas, restablecen el sentido transmitido por los textos y están relacionadas con el proceso de comprensión y reexpresión del sentido (Hurtado 2011, 218).

Equivalencia funcional y comunicativa

A partir de los años ochenta y noventa se introduce en la descripción de la equivalencia traductora el aspecto funcional y comunicativo, con lo que se consolida el carácter contextual de la definición. Autores como Neubert (1985), De Beaugrande y Dressler (1981) resaltan el hecho de que la equivalencia solo puede darse entre textos.

Reiss y Vermeer (1996) introducen el concepto de función y proponen una diferenciación entre equivalencia y adecuación. Estos autores consideran la adecuación como el criterio dominante. Mientras que en otras teorías la adecuación es una característica del texto meta en relación a la actividad traductora, en la teoría del *skopos* la adecuación es tan importante que define la equivalencia como la adecuación al *skopo*. Para que se dé la equivalencia, la traducción debe cumplir la misma función comunicativa que en el texto original y de esta por ello podemos afirmar que se postula un concepto de adecuación que, además de ser dinámico, engloba al de equivalencia.

1.1.2 La finalidad de la traducción

Viendo todos los problemas e incertidumbres que arroja el concentrarse en el concepto de equivalencia como parámetro básico para definir el proceso de traducción aparecen a partir de los años ochenta otras teorías que no solo cuestionan el criterio de la equivalencia, sino que además se concentran en otra perspectiva. Pym (2012, 47) las define como teorías de la finalidad, pero se está refiriendo directamente a la teoría del *skopo* de Reiss y Vermeer (1996). El término *skopos* = *Skopus* es introducido por Hans J. Vermeer como término técnico para definir el propósito de una traducción y de la acción traductora.

Si partimos de la premisa de que una acción, incluso la traducción como una acción de comunicación, está determinada por un propósito, el propósito de un acto comunicativo, su *skopo* se satisface si el resultado obtenido cumple la intención del agente (del emisor) y si el

1. Traducción, traducción técnica, problemas de la traducción

destinatario de la acción comunicativa en su propia situación puede interpretar los mensajes recibidos de manera concluyente. Como *skopo* se denomina el alcance efectivo y real de las palabras que determina o define una expresión de un modo particular. El *skopo* de una traducción es determinado por la función del *translats* o texto traducido (Nord 1991, 27).

La teoría del *skopo* asume que el propósito de un *Translats* de cumplir una función particular, es el factor determinante al que se debe ajustar el proceso de traducción. Es decir, una traducción puede en principio satisfacer cualquier función de un texto, la evaluación de una función en particular como “buena”, “útil” o “ética” se deja de lado. La calidad de una traducción solo se evaluará en términos del cumplimiento de su función particular.

La teoría del *skopo* le da un papel primordial al tema de la función y de la posibilidad de cambiar esa función en el texto meta. El tipo de encargo de traducción determina la estrategia translativa que se debe seguir y si se debe mantener o cambiar la función del texto. Si la traducción o texto meta debe cumplir funciones comunicativas diferentes de las del texto original, los textos ya no tendrán la misma relación y no se podría hablar de equivalencia, sino que se tendría que hablar de adecuación. La adecuación se definiría como:

- La conformidad de la elección de signos en la lengua meta con respecto a la dimensión seleccionada en el texto original.
- Se trata de adecuar la traducción a la finalidad que con ella se pretende, escogiendo los signos apropiados a esa finalidad.
- La relación que existe entre el texto meta y el texto original teniendo en cuenta el *skopo* que persigue.

Reiss y Vermeer proponen pues un concepto de adecuación que abarca el de equivalencia: una traducción es adecuada cuando la elección de los signos en el idioma meta depende de la finalidad de la traducción y una traducción es equivalente cuando la función del texto original y la del texto meta considere conveniente que permanezcan invariables. En este sentido cambian la perspectiva postulada por la lingüística estructuralista (Edward Sapir, Benjamin Whorf y Ferdinand de Saussure), según la cual es prácticamente que imposible encontrar

1. Traducción, traducción técnica, problemas de la traducción

una equivalencia entre las lenguas, ya que cada idioma tiene diferentes estructuras y sistemas (Pym 2012, 23–27).

Con equivalencia y finalidad tenemos dos de los conceptos que aparecen repetidamente y que son claves para el análisis de la traducción. Sin embargo no son estos los únicos conceptos importantes. Hurtado lista junto con la equivalencia otras nociones como la invariabilidad traductora, la unidad de traducción, el método traductor, las técnicas de traducción, las estrategias traductoras, los problemas y los errores de traducción” (Hurtado 2011, 201).

El tema central de nuestra tesis son precisamente dos de estas nociones fundamentales. Nos referimos a los problemas y los errores de traducción. En el capítulo “Problemas de la traducción” nos dedicaremos en un primer momento a la definición general de estos conceptos, para luego ver si se presentan problemas o errores especiales en la combinación lingüística alemán-castellano y si son estos problemas y errores u otros distintos los que nos encontramos en la traducción automática. Pero antes de empezar con los problemas y errores de la traducción introduciremos otro tema central de nuestra tesis: la traducción de textos técnicos.

1.2 La traducción técnica

La traducción técnica puede situarse dentro de la traducción especializada (Gamero 2001, 23). De esta manera se está indicando que este tipo de traducción, como la científica o la jurídica, es diferente de la traducción literaria o audiovisual. Sin embargo esta denominación no es ajena a controversias porque en cierta medida otros tipos de traducciones también son traducciones especializadas.

Al definir un tipo de traducción como ‘especializada’ no se está haciendo referencia al proceso de traducción, sino al tipo de texto traducido, ya que en el texto se utiliza un lenguaje de especialidad: lenguaje técnico, científico, jurídico, económico etc. que obliga a los traductores a tener un cierto conocimiento especializado y a dominar la terminología correspondiente (Gamero 2001, 23).

Hurtado propone una clasificación de la traducción partiendo de una variable que denomina modalidades de traducción o modo traductor. Como modo traductor define Hurtado “...la variación que se produce en la traducción según las características del modo del texto

1. Traducción, traducción técnica, problemas de la traducción

original y de la traducción” (2001, 69). Las modalidades (principales) de traducción según este criterio sería: traducción escrita, traducción a la vista (oral), interpretación simultánea, interpretación consecutiva, interpretación de enlace, susurrando (simultánea al oído del destinatario), doblaje, voces superpuestas, subtitulación, traducción de programas informáticos, traducción de productos informáticos multimedia, traducción de canciones, supratitulación musical y traducción iconográfica. Estas modalidades no se excluyen y pueden darse paralelamente, por ejemplo al realizar una la interpretación simultánea de una película; también puede empezarse una traducción con una modalidad y cambiar a otra (una interpretación simultánea, que luego pasa a ser una interpretación de enlace).

Los tipos de traducción tienen que ver según Hurtado con la traducción de textos que pertenecen a determinados ámbitos socioprofesionales. En la traducción escrita los tipos de traducción se relacionan con las áreas convencionales tradicionales, aquí hablamos de traducción literaria, general o especializada. La traducción técnica pertenece a esta última categoría. Es decir, a la traducción de textos especializados o géneros especializados. Es especializada porque requiere de conocimientos y habilidades especiales.

También tenemos que especificar que nos estamos refiriendo a la traducción técnica y no a la traducción científica. Los textos científicos tienen la función de difundir un conocimiento, mientras que los textos técnicos no se escriben para difundir conocimiento, sino para cumplir una serie de requisitos legales o hacer publicidad a productos (Pinchuck 1977). Otros autores como Jumplet (1961) afirman que el lenguaje científico es más universal que el técnico y que hay más paralelismo en las lenguas involucradas que en el de la técnica. Gamero afirma que en los textos científicos predomina la argumentación y la descripción, mientras que en los textos técnicos dominan el carácter descriptivo y exhortativo (Gamero 1998).

En sentido comunicativo los textos técnicos pueden utilizarse (como los científicos) para transmitir el conjunto de saberes propios de una disciplina a especialistas en formación o en algunos casos para divulgar unos conocimientos básicos entre el público general. El ámbito del uso de los textos técnicos es más amplio que el de los científicos: se pueden producir textos que contribuyan a la organización de procesos industriales, que ofrezcan información al usuario de productos, que anuncien productos (Gamero 2001, 28).

1. Traducción, traducción técnica, problemas de la traducción

1.2.1 *Definición del texto técnico*

Gamero define el texto técnico como “un acto concreto de comunicación en el que los emisores son ingenieros, técnicos o profesionales; los receptores son otros ingenieros, técnicos, especialistas en formación o el público en general; la situación comunicativa está relacionada con la industria, la explotación agrícola, la fabricación de productos o la oferta de servicios; el foco predominante es la exposición o la exhortación; el modo es generalmente escrito; el campo es de carácter exclusivamente técnico...” Aquí se refiere la autora a los epígrafes 31 y 33 de la nomenclatura internacional de la UNESCO sobre la clasificación de los campos de la técnica (Gamero 2010, 38). También resalta la autora que en este tipo de texto no se utilizan dialectos temporales, geográficos ni idiolectos y que sus rasgos intratextuales son variados y están determinados por las convenciones del género. Como elementos intratextuales que definen los textos técnicos clasifica la autora en el plano léxico la terminología y la utilización de siglas y símbolos; en el plano morfológico el uso de determinados tiempos verbales, conjunciones y pronombres; en el plano sintáctico los tipos de oración empleados y en el plano textual la progresión textual, los recursos de coherencia y cohesión.

Los autores Sevilla y Muñoz (2003) hacen una distinción de tres niveles de comunicación del texto científico-técnico:

1. Comunicación entre especialistas a través de publicaciones (escritas) o conferencias (oral). El contexto será formal y el discurso culto.
2. Comunicación entre científicos y técnicos en un contexto informal a través de diarios de laboratorio (escrito) o discusiones informales (oral).
3. La transmisión de información científica y técnica al público en general en publicaciones divulgativas (escrito) o programas divulgativos de radio o televisión (oral). Aquí podríamos incluir también a través de revistas o foros en Internet. El emisor no necesariamente tendrá que ser un especialista y la comunicación puede ser tanto formal como informal, dependiendo del público.

Hoffmann hace una distinción de 5 niveles de comunicación relacionados con sus respectivos agentes modelo con el que se explican las diferencias entre el lenguaje común y los lenguajes *especializados*, y además las diferencias entre diferentes registros dentro de los len-

1. Traducción, traducción técnica, problemas de la traducción

guajes utilizados para la comunicación profesional (Hoffmann 1987, 64). El modelo ofrece también indicaciones que se refieren al léxico o terminología y propone una clasificación ‘vertical’ del lenguaje especializado, que tiene en cuenta el nivel de abstracción, la forma lingüística, el ámbito y los participantes (1987, 64–71).

	Nivel de abstracción	Forma lingüística	Ámbito	Participantes
A	Más elevado	Símbolos artificiales para elementos y relaciones	Ciencias fundamentales teóricas	Científico - técnico
B	Muy elevado	Símbolos artificiales para elementos; lenguaje general para relaciones (sintaxis)	Ciencias experimentales	Científico (técnico) – científico (técnico)
C	Elevado	Lenguaje natural con terminología especializada y sintaxis controlada	Ciencias aplicadas y técnicas	Científico (técnico) – directores científico-técnicos de la producción material
D	Bajo	Lenguaje natural con terminología especializada y sintaxis relativamente libre	Producción material	Directores científico-técnicos en la producción material – maestros trabajadores, especialistas
E	Muy bajo	Lenguaje natural con algunos términos especializados y sintaxis libre	Consumo	Directores de producción, comerciantes, consumidores

Tabla 1: Niveles del lenguaje especializado según Hoffmann

Un lenguaje puede pero no tiene que tener todos estos niveles. Algunas disciplinas no llegan al nivel más alto de abstracción (A). Los niveles se mezclan en actos de comunicación concretos, el matemático necesita del lenguaje natural (nivel C) para explicar sus fórmulas que pertenecen al nivel más alto de abstracción. El médico tiene que partir del nivel C para comunicarse con sus pacientes que se encuentran en nivel D, e incluso en el nivel D utilizamos fórmulas, números y signos que pertenecen al nivel de abstracción A.

Hoffmann propone también una distribución de niveles de especialización a nivel horizontal (1987, 58–61), en la que se clasifican los diferentes temas en una secuencia que refle-

1. Traducción, traducción técnica, problemas de la traducción

ja la relación lingüística de los lenguajes especializados a nivel léxico y sintáctico. Dicha clasificación no pretende ser completa empieza con el lenguaje de las humanidades, continúa con el lenguaje técnico y termina con el lenguaje científico. En este sentido los lenguajes técnicos o especializados se pueden dividir en dos sentidos: 1) por criterios puramente técnicos y 2) de acuerdo con el contenido y el punto de vista lingüístico.

1.2.2 Características específicas de la traducción técnica

Después de ocuparnos de la definición del tipo de texto técnico, trataremos de determinar qué entendemos por la traducción técnica. Gamero (2001) afirma que las características de la traducción técnica están en parte determinadas por el texto de partida —o sea por el texto técnico— y por las competencias que debe tener el traductor. Por ello establece dos tipos de categorías, aquellas que están relacionadas con el texto de partida: el campo textual, la terminología y el género textual; y las que están relacionadas con la competencia traductora:

Tipo de categoría	Características
Categorías relacionadas con el texto de partida	Campo textual, aquí se incluyen los campos clasificados por la UNESCO como ciencias tecnológicas (Gamero 2001, 27) e incluye también las ciencias agrarias
	Terminología específica
	Género textual
Categorías relacionadas con la competencia del traductor:	Los textos técnicos obligan al traductor a tener conocimientos en el campo temático
	El traductor debe utilizar la terminología correcta
	Debe respetar las características de los géneros correspondientes
	Debe dominar la documentación como herramienta de trabajo (conocer las obras de referencia disponibles)

Tabla 2: Categorías de la traducción técnica según Gamero

Gamero define el género técnico como “un prototipo de texto, utilizado en determinadas situaciones comunicativas que se repiten dentro de una cultura concreta en cualquiera de los ámbitos incluidos en la nomenclatura de la UNESCO para los campos de la tecnología, y

1. Traducción, traducción técnica, problemas de la traducción

que tiene la finalidad de que la comunicación se efectúe del modo más eficaz posible” (2010, 62).

Como géneros técnicos define el manual de instrucciones general, el manual de instrucciones especializado, el acta de reunión técnica, etc. Para la clasificación de los textos se basa la autora primero en una clasificación de los géneros técnicos según su función textual en géneros expositivos, expositivos con foco secundario exhortativo y exhortativos con foco secundario expositivo. A esta clasificación añade Gamero el papel del receptor, es decir si éste tiene conocimientos generales o especializados sobre el tema tratado. Teniendo en cuenta estos aspectos establece la siguiente clasificación:

Foco contextual	Tipo de receptor	
	General	Especializado
Expositivo	➤ Artículo divulgativo	➤ Acta de reunión técnica
	➤ Monografía divulgativa	➤ Descripción técnica
	➤ Enciclopedia técnica	➤ Informe técnico
Expositivo con foco secundario exhortativo		➤ Listado de piezas
		➤ Manual técnico
	➤ Artículo comercial	➤ Anuncio en medio especializado
	➤ Folleto informativo publicitario	➤ Comunicación interna de empresa
Exhortativo	➤ Memoria anual	
	➤ Manual de instrucciones general	➤ Manual de instrucciones especializado
		➤ Instrucciones de trabajo
		➤ Plan de producción
		➤ Plan de estudios
		➤ Patente
		➤ Norma técnica
	➤ Norma laboral	
	➤ Certificado técnico	

	1. Traducción, traducción técnica, problemas de la traducción	
Exhortativo con foco secundario expositivo	➤ Anuncio técnico en medio general	➤ Carta técnica
	➤ Folleto publicitario informativo	➤ Pliego de condiciones
	➤ Publireportaje	➤ Proyecto técnico
	➤ Prospecto de medicamento	➤ Solicitud de desarrollo del producto

Tabla 3: Clasificación de la traducción técnica escrita de Gamero (2001, 69)

La norma europea EN 82079 de septiembre 2012 regula el diseño y preparación de manuales de instrucciones para máquinas, instalaciones y productos técnicos. Esta norma reglamenta la manera en que deben estar estructurados los manuales, el contenido y presentación tanto de productos simples como complejos. Esta nueva norma sustituyó la norma EN 62079 de abril 2001. También existen un gran número de leyes, directrices y normas que reglamentan qué datos deben estar incluidos, en función de las características del producto y del área de aplicación. La edición nacional de Alemania apareció como norma DIN EN 82079-1 VDE 0039-1 se publicó en junio de 2013.

Como indicamos en la introducción los segmentos que hacen parte de nuestro corpus son traducciones automáticas de textos técnicos, es decir según la clasificación de Gamero son textos con un foco textual exhortativo general. Debido a que son traducciones automáticas y no de traductores humanos en nuestro marco teórico no nos ocuparemos de las categorías relacionadas con la competencia del traductor.

1.3 Problemas de la traducción

Para poder abordar el problema central de esta tesis, es decir, el análisis de los problemas que se pueden presentar en la traducción automática, necesitaremos una clasificación general, que no se guíe en las habilidades y competencias del traductor o poseedor, sino en las características mismas del proceso de traducción y de los idiomas involucrados en este proceso. Muchos autores y traductólogos han realizado clasificaciones de los problemas de traducción. Aquí daremos una vista general de algunas de las definiciones que consideramos de utilidad para nuestra tesis.

1. Traducción, traducción técnica, problemas de la traducción

1.3.1 Los problemas de la traducción según Nord

Según la visión funcionalista de la traducción, especialmente de Nord, en todo proceso de traducción se enfrentan dos situaciones: la situación original o base, en la que se emplea un texto a través del que se establece un tipo de comunicación entre un emisor y un destinatario en la misma cultura original, y la situación meta en la que se utilizará el texto traducido (meta), para también establecer una comunicación, ahora entre el autor del texto origen y un receptor en la cultura meta (Nord 2009, 234). En este proceso suelen presentarse diferentes tipos de problemas que el traductor debe solucionar para que se pueda dar el proceso comunicativo. Nord se refiere a problemas —no a dificultades subjetivas— ya que los problemas no dependen de la competencia del traductor, ni de las condiciones técnicas, sino que se encuentran tanto en el texto original, como en las diferencias estructurales de los idiomas y en la distancia entre las culturas de los idiomas implicados (Nord 1987, 5–8). Los problemas, a diferencia de las dificultades, son “intersubjetivos, generales, y han de ser solucionados mediante procedimientos traslativos que forman parte de la competencia traductora” (Nord 2009: 233).

Según Nord (2009) podemos hablar de diferentes tipos de problemas de traducción: (a) problemas pragmáticos, (b) problemas culturales o relacionados con convenciones culturales, (c) problemas lingüísticos o (d) problemas de traducción específicos del texto.

En una primera aproximación Nord habla solo de problemas pragmáticos, culturales y de problemas específicos del par de idiomas envueltos en el proceso de traducción (Nord 1991, 182). En escritos y presentaciones posteriores la autora añade una clase más, la de los problemas propios del texto (a traducir) y especifica una de las categorías mencionadas. Se referirá concretamente no a los problemas de las lenguas involucradas, sino a problemas lingüísticos de la traducción (Nord 2009, 235–236). A continuación haremos una breve reseña de la clasificación de problemas de Nord.

1.3.1.1 Los problemas pragmáticos

Los problemas pragmáticos tienen que ver con la situación comunicativa y surgen de la diferencia entre el texto original y la situación meta. Por ejemplo, cuando un texto es escrito para un público especial, los habitantes de una región o los hablantes de un idioma, y el mismo texto tiene que ser traducido a otro idioma, para lectores que viven en regiones o con-

1. Traducción, traducción técnica, problemas de la traducción

diciones diferentes, en muchos casos el traductor tendrá que ajustar el texto original, así esto implique no ser fiel a este (Nord 2001a, 16). Este tipo de problemas también se dan cuando el autor del texto original presupone en sus receptores determinados conocimientos propios del entorno de su cultura, que no necesariamente tienen los lectores o receptores del texto meta (Nord 2009, 234). También datos temporales exactos, como estaciones o fechas que sitúen cronológicamente o geográficamente el texto original, tendrán que ser ajustadas en el texto traducido.

Según esta autora los problemas pragmáticos pueden identificarse si se contraponen los factores externos del texto en la situación original y en la situación meta; los problemas pueden surgir tanto por la diferencia, como por la similitud de los factores externos (Nord 2001a, 150). En la tabla siguiente resumiremos los factores externos que define la autora.

Factor	Descripción
Referencia al emisor o productor del texto	Nombre del autor o productor del texto, de instituciones públicas o empresas, formas verbales o pronominales, paráfrasis, lenguaje empleado para caracterizarlo, la perspectiva que se le asigna, la manera en que se califica.
Intención del autor	Apelación indirecta al receptor por medio de adjetivos u otros recursos que le dan o quitan valor y peso al texto.
Referencia al receptor	Nombre, segunda persona, formas de cortesía, perspectiva del receptor, su trasfondo cultural.
Referencia o influjo del medio	Espacio, utilización de varios idiomas, indicaciones de fuentes.
Referencias locales, temporales, motivación	Aquí, a mi lado, yo, nosotros, hoy, este año.
Función textual	Ejemplos, comparaciones, citas, subtítulos.

Tabla 4: Factores externos que pueden generar problemas pragmáticos según Nord (2001a, 151-68)

1.3.1.2 Problemas culturales de la traducción

Nord enfatiza que en cada proceso de traducción están involucrados dos sistemas culturales con sus respectivas convenciones. En algunos escritos la autora se refiere a esta cate-

1. Traducción, traducción técnica, problemas de la traducción

goría como problemas en el uso de las convenciones (2001b, 15), porque las convenciones si bien facilitan la interacción, son aleatorias y están determinadas por la cultura. Los problemas culturales se originan precisamente en la diferenciación entre las normas, las convenciones verbales o no verbales y los hábitos específicos de una cultura (por ejemplo, las convenciones de género, las convenciones generales de estilo, las convenciones de medida, etc.). El traductor debe decidir si adapta o no las convenciones de la cultura base a la cultura meta; las convenciones deben respetarse y transcribirse si se trata de una traducción documento más, no sé si trata de una traducción instrumento. La traducción-documento tiene según Nord la función de producir un documento sobre una acción comunicativa en la cultura base para los lectores de la cultura meta. Es decir, se reproduce el texto, el contenido e incluso la forma del texto origen, sin hacer prácticamente ninguna adaptación a la cultura o situación meta. Por otra parte la traducción-instrumento pretende crear una nueva acción comunicativa entre el emisor de la cultura base y el público de la cultura meta adaptando la información recibida en el texto origen a las condiciones de la cultura meta (Nord 2009, 226-31).

Existen convenciones de estilo que afectan las áreas de la terminología léxica, la sintáctica y la construcción textual y se hacen relevantes para la traducción, si difieren en los idiomas en cuestión. Nord distingue diferentes tipos de convenciones: convenciones generales, convenciones literarias, convenciones formales y convenciones en la escritura y traducción de nombres propios.

1) Entre las convenciones generales de estilo nos encontramos con las convenciones léxicas, como por ejemplo la personificación (La casa blanca informó que el presidente estaría ausente). También encontramos —especialmente en textos publicitarios— la utilización de elementos como conjunciones en un sentido distinto del habitual e incluso en un sentido completamente diferente. También habla Nord de convenciones generales sintácticas como la diferencia entre estilo nominal y verbal, las elipses, aposiciones, oraciones de relativas pueden generar problemas en la traducción (Nord 2001a, 182).

2) La autora también habla de convenciones literarias como aquellos medios estilísticos literarios que se utilizan en textos no literarios, como por ejemplo en textos publicitarios, para producir un efecto determinado y que deben ser adaptados a la situación y cultura meta.

1. Traducción, traducción técnica, problemas de la traducción

Hablamos aquí —por ejemplo— de epítetos que se utilizan para describir un paisaje en un prospecto turístico. “Verdes valles, verdes lomas, bosques de alma verde, Asturias es la consagración eterna de lo verde, el camino para huir de lo cotidiano” (Nord 2001a, 128).

3) Existen además convenciones formales que se utilizan para caracterizar y marcar determinados elementos textuales, como la caracterización tipográfica de ciertos elementos, la utilización de extranjerismos o el escribir nombres propios en mayúsculas. También a este grupo pertenece la manera de escribir números, medidas, fechas y abreviaturas. Las reglas de puntuación hacen parte de este grupo de convenciones que pueden variar dependiendo del idioma y cultura meta.

4) Otro tipo de convenciones que difieren según la cultura son aquellas que tienen que ver con la adaptación, traducción o transcripción de nombres propios. En algunos idiomas no suelen traducirse los nombres propios, de los reyes. En alemán se habla de “König Juan Carlos” o “Königin *Elisabeth* von England”, mientras que en castellano hablamos de “La reina *Isabel* de Inglaterra” (Nord 2001a, 190).

1.3.1.3 Problemas lingüísticos de la traducción

En cada proceso de traducción se enfrentan dos sistemas lingüísticos con características particulares. Las características propias de cada sistema pueden ser diferentes y por ello generar problemas de traducción. Los traductores tendrán que buscar soluciones especiales para poder transmitir en el texto meta lo que el autor quería transmitir al haber escogido este tipo de formulación.

1.3.1.4 Problemas específicos del texto

Bajo esta categoría agrupa Nord los problemas con que se enfrenta el traductor al tratar de traducir ciertas figuras estilísticas, juegos de palabras, neologismos, metáforas e incluso defectos lingüísticos, que el autor del texto origen ha empleado para lograr ciertos efectos o funciones comunicativas. Este tipo de problemas no se presentan en textos convencionales o estandarizados, sino en textos literarios o publicitarios, donde el autor tiene una libertad mayor. En algunos artículos denomina Nord esta clase de tipo de **problemas de traducción extraordinarios** (Nord 2001a, 236) porque la solución que el traductor encuentra para resolverlos no podrá utilizarse en otros problemas del mismo tipo.

1. Traducción, traducción técnica, problemas de la traducción

En la tabla siguiente resumiremos los tipos de problemas que Nord clasifica y una definición breve de estos.

Tipo de problema	Definición
Pragmáticos	Tienen que ver con la situación comunicativa y surgen de la diferencia entre el texto original o base y la situación meta.
Culturales /convencionales	Se originan en la diferenciación entre las normas, las convenciones verbales o no verbales y los hábitos específicos de cada cultura, por ejemplo convenciones de género, dialectos, léxico de una época.
Lingüísticos	Problemas que se dan cuando se enfrentan dos sistemas lingüísticos con características particulares e incluso diferentes.
Específicos del texto	Problemas con que se enfrenta el traductor al tratar de traducir ciertas figuras estilísticas que el autor del texto origen ha empleado para lograr ciertos efectos comunicativos.

Tabla 5: Clasificación de los problemas de traducción según Nord

1.3.2 *Los problemas de la traducción según Krings*

Existen otras clasificaciones de problemas de traducción que se basan –entre otros aspectos– en el estudio de la competencia que el traductor debe desarrollar para tomar decisiones cuando tiene que resolver los problemas que se presentan en el proceso de traducción. Según el modelo de competencia traductora, la resolución de problemas sería una especie de competencia estratégica subordinada, que comprende no solo el desarrollo de la capacidad de identificar los problemas, sino también la capacidad de buscar y evaluar diferentes soluciones (Kelly 2002, 18). Siguiendo esta idea Krings realizó una investigación con un grupo de traductores profesionales utilizando el método de protocolos de verbalización (o *TAP*, del inglés *Think-Aloud Protocols*), para tener una aproximación a los procesos de toma de decisiones que ocurren en la mente de los traductores durante la traducción. A partir de este estudio empírico Krings hace una diferenciación entre los problemas que tienen que ver con la comprensión del texto origen (problemas de recepción) y los que tienen que ver con la aplicación de lo que se entiende en el idioma meta (problemas de reproducción). Krings también encuentra en su estudio que se dan problemas que involucran ambos aspectos, es decir cuya causa se encuentra tanto en la recepción o comprensión del texto origen, como en la de reproducción de

1. Traducción, traducción técnica, problemas de la traducción

lo comprendido. Esta forma combinada la denomina el autor *reception-reproduction problem* (Krings 1986, 146).

Tipos de problemas	• Problemas que tienen que ver con la comprensión del texto origen
	• Problemas en la reproducción de lo comprendido
	• Problemas que involucran ambos aspectos: comprensión/reproducción

Tabla 6: Tipos de problemas según Krings

En un artículo posterior Krings formula otra categorización de problemas, esta vez dependiendo del nivel en que se presenten, es decir a nivel de los lexemas, a nivel de los morfemas o palabras, a nivel de varias palabras, a nivel de partes de las oraciones, a nivel de la oración, a nivel del texto o a nivel metatextual (Krings 2005, 354). Esta clasificación se asemeja a la definición que Nord hace de problemas de carácter lingüístico. Los problemas que Krings define como problemas de comprensión y/o reproducción, pueden ser ocasionados porque se presentan algún tipo de problemas pragmáticos, culturales o convencionales o lingüísticos.

Nivel en que se dan los problemas	• Lexemas-monemas / palabras / combinaciones de palabras
	• Frase / combinaciones de frases
	• Texto /Meta-texto

Tabla 7: Clasificación de los problemas de Krings según el nivel en que se presentan

1.3.3 Los problemas de la traducción según Schmitt

Siguiendo también una orientación funcionalista Schmitt define los problemas de la traducción apoyándose en la definición de Wilss (1977, 202) según la cual un problema aparece siempre que no se de una correspondencia exacta de tipo léxico, semántico o sintáctico entre los segmentos del texto origen y la traducción y donde los procedimientos de traducción sustitutivos claramente ocasionan disparates lingüísticos. Schmitt clasifica los problemas de traducción en el contexto de la traducción de textos técnicos. Al igual que Krings considera como problemas de traducción aquellos que pueden surgir en la comprensión del texto y clasifica los problemas en dos grupos: evidentes u obvios y latentes u ocultos.

1.3.3.1 Problemas evidentes

Problemas evidentes son aquellos que se dan porque hay un déficit en el manejo del idioma (tanto original como de traducción), defectos en el texto original o déficit de conocimiento en la materia del texto, es decir, conocimiento especializado (Schmitt 1999, 57–58). Este tipo de problemas no constituyen en realidad un problema para el proceso comunicativo, porque son tan evidentes, que generan una especie de interrupción en el proceso de comunicación, que únicamente se podrá dar cuando éstos sean resueltos. La comunicación se deteriora en la medida en que no se entiende lo que se quiere comunicar, pero mucho más problemático sería no reconocer el problema y creer erróneamente que se entiende lo que se trata de comunicar. Problemas de este tipo ocurren cuando hay un déficit en el idioma original, en el de traducción o en ambos. Por ejemplo, cuando por desconocimiento del significado de ciertos conceptos se presentan lagunas léxicas en la traducción. Este tipo de lagunas no significan necesariamente que el traductor no maneje bien el idioma, en algunos casos se dan estos problemas porque el traductor escoge la estrategia errónea para su solución. También se presentan cuando hay defectos en el texto que se va traducir o un déficit en el conocimiento en la materia. Defectos en el texto original se dan cuando el texto no está completo, es ambiguo o tiene algún tipo de incoherencia. Un déficit en el conocimiento especializado se da cuando en el texto de partida no se entiende el significado de una palabra o de ciertas expresiones. Incluso un buen traductor puede tener este tipo de déficit. En el texto original, el traductor se ve siempre confrontado con la ambigüedad y la vaguedad de ciertas las expresiones lingüísticas. Schmitt nos pone como ejemplo de la expresión *power tilt* cuya significación (barra colectora de energía azimutal de un reactor nuclear) solo es clara para un público especializado (Schmitt 1999, 58).

1.3.3.2 Problemas latentes

Los problemas latentes surgen según Schmitt cuando se dan incongruencias culturales. También aquí establece el autor diferentes categorías que dependen del tipo de incongruencia que se dé, es decir incongruencia funcional, formal o semántica. Incongruencias funcionales se dan cuando un tipo de texto en la cultura original tiene otra función en la cultura meta. Schmitt también hace una diferenciación entre las incongruencias semánticas como denotativas o connotativas.

1. Traducción, traducción técnica, problemas de la traducción

Tipo de problemas	Características	
Problemas evidentes	Déficit en el texto de partida o en el texto traducido	Se presentan cuando se desconoce el significado de un concepto o expresión en el texto original y/o la correspondencia en el idioma meta.
	Defectos en el texto de partida	Defectos formales, errores en medidas o cifras, errores ortotipográficos, discrepancias entre el texto y las imágenes correspondientes o entre el texto y la realidad, falta de claridad en diferentes niveles, errores de expresión, errores en el contenido.
	Déficit de conocimiento de especializado	Cuando en el texto de partida a nivel de la <i>parole</i> no se conoce la significación de un concepto o expresión.
Problemas latentes	Incongruencias funcionales	Cuando la misma función comunicativa en la cultura origen y meta se puede cumplir con tipos de textos diferentes.
	Incongruencias formales (diferencias culturales)	Externas, es decir la apariencia de un texto.
		Internas en sentido formal, como las características estilísticas y sintácticas de un texto.
	Incongruencias semánticas	denotativas
connotativas		

Tabla 8: Clasificación de los problemas de la traducción según Schmitt

1.3.4 Los problemas de la traducción según Hurtado

Amparo Hurtado define los problemas de traducción teniendo en cuenta la propuesta de Nord como “las dificultades (lingüísticas, extralingüísticas, etc.) de carácter objetivo con que puede encontrarse el traductor a la hora de realizar una tarea traductora” (Hurtado 2011,

1. Traducción, traducción técnica, problemas de la traducción

286). La autora identifica cinco categorías básicas de problemas que son también las planteadas por el “modelo holístico de PACTE”: problemas lingüísticos, textuales, extralingüísticos, problemas de intencionalidad y pragmáticos (2011, 288).

1. Los **problemas lingüísticos** son para la autora aquellos “problemas relacionados con el código lingüístico fundamentalmente en el plano léxico (léxico no especializado) y morfológico” (2011, 288). Estos problemas se derivan especialmente de las discrepancias entre las lenguas y pueden ser de comprensión y/o de reexpresión.
2. Como **problemas textuales** define Hurtado los problemas que remiten a cuestiones de coherencia, progresión temática, cohesión, tipologías textuales (convenciones de género) y estilo. Se originan en las diferencias de funcionamiento textual de las lenguas y también pueden ser de comprensión y/o de reexpresión (2011, 288).
3. Los **problemas extralingüísticos** de acuerdo con Hurtado son problemas “remiten a cuestiones temáticas (conceptos especializados), enciclopédicas y culturales”(2011, 288). Es decir, se relacionan con cuestiones culturales.
4. **Problemas de intencionalidad** son los problemas relacionados con las dificultades de comprensión de la información del texto original (intención, intertextualidad, actos de habla, presuposiciones, implicaturas).
5. **Problemas pragmáticos** son aquellos “que se derivan del encargo de traducción, de las características del destinatario y del contexto en que se efectúa la traducción. Afectan la reformulación” (2011, 288).

Hurtado, a diferencia de Nord, hace un análisis más global de los problemas, ya que para esta autora se deben primero realizar estudios empíricos en los diferentes tipos y modalidades de traducción que ayuden a identificar y clasificar los problemas a que se enfrenta el traductor, así como sus reacciones y las acciones o estrategias que éste desarrolla para resolver los problemas encontrados. También considera importante que se realicen estudios empíricos que analicen el proceso de adquisición de la competencia traductora, es decir, los problemas con que se encuentra el traductor durante su formación.

1. Traducción, traducción técnica, problemas de la traducción

Tipo de problemas	Características
Lingüísticos	Discrepancias entre las dos lenguas en diferentes planos (léxico y morfológico). Se presentan tanto en la comprensión y/o reexpresión.
Textuales	Problemas de coherencia, progresión temática, cohesión, convenciones de género y estilo. Se presentan tanto en la comprensión y/o reexpresión.
Extralingüísticos	Concernientes a la temática a cuestiones enciclopédicas y culturales.
Intencionales	Relacionados con las dificultades de comprensión de la información del texto original
Pragmáticos	Problemas derivados del encargo de traducción, de las características del destinatario y del contexto en que se efectúa la traducción. Afectan la reformulación.

Tabla 9: Clasificación de los problemas de traducción según Hurtado

La clasificación de Hurtado es similar a la de Nord, si bien podríamos afirmar que los problemas que la autora define como extralingüísticos se encuentran en la definición de los problemas culturales de Nord. Los problemas que Hurtado clasifica como textuales son muy similares a los que se encuentran en la clasificación de Nord de problemas de convenciones.

1.4 Errores en la traducción

Los errores en la traducción se producen cuando no se solucionan correctamente los problemas que se presentan en el proceso de traducción. También nos percatamos de la existencia de los problemas de la traducción cuando vemos los errores que se dan, es decir, cuando no se solucionan. Por ello la temática de los problemas que se presentan en el proceso de la traducción no se puede contemplar sin tener en cuenta los errores que se surgen cuando se ha concluido dicho proceso. A las categorías de problemas que hemos presentado anteriormente corresponde también una categorización de tipos de errores. A partir de la clasificación que daremos a continuación podremos establecer más adelante qué tipos de errores de traducción consideramos relevantes para la evaluación de la traducción de los textos que aparecen en el corpus.

1. Traducción, traducción técnica, problemas de la traducción

1.4.1 La clasificación de errores de traducción según Nord

La meta del proceso de traducción está determinada por el modelo desde el que se interprete. Si se ve el proceso de traducción como una equivalencia, el texto original será el modelo a reproducir. Si vemos el proceso de traducción desde la perspectiva funcionalista es el *skopo* el que define la meta. Es decir, dependiendo de la perspectiva se puede dar una definición de los tipos de errores diferentes.

Según el funcionalismo una traducción es adecuada si “logra las funciones comunicativas pretendidas por el cliente o iniciador que la ha encargado” (Nord 2009, 237). Es decir el que se reproduzca correctamente en el idioma meta la finalidad¹ con que ha sido redactado, producido o encargado el texto, determina si es *correcta* o no.

Si solo se quiere entender de qué se trata el texto, sin importar si la traducción es correcta gramatical o culturalmente, la traducción ya estaría cumpliendo su función y sería correcta. Siguiendo la teoría funcionalista de Nord (2009, 238–239) nos encontramos con los siguientes tipos errores de traducción:

Tipo de error	Definición
Errores pragmáticos	Aquellos errores que si no se corrigen pueden ocasionar que la función comunicativa del texto original no se cumpla.
Errores culturales o convencionales	Son aquellos errores que no destruyen la comunicación, pero pueden dar una luz negativa sobre el autor o el traductor.
Errores lingüísticos	Son errores específicos del par de idiomas involucrados.
Errores específicos del texto	Son desviaciones del estándar del texto meta, que no son errores de traducción, sino que son ocasionados —por ejemplo— por la falta de dominio de las convenciones estilísticas y de género del idioma meta.

Tabla 10: Tipología de errores de Nord

¹ El concepto de finalidad se basa en la teoría de los actos de habla, desarrollada entre otros por Searle (1969) y también tematizada por la lingüística pragmática. Según esta corriente un acto de habla tiene varios niveles fundamentales: cada vez que se expresa algo, no solo se dice algo con palabras (acto locutivo), sino que se realiza algo al decir (acto ilocutivo), se pregunta, se afirma, se da una orden; además de expresar lo dicho, de la manera en que se ha dicho, también existe una intención no expresa (acto perlocutivo): por ejemplo, yo puedo pretender ofender a una persona al hacer una pregunta.

1.4.2 La clasificación de los errores de traducción según Hurtado

Para Hurtado la noción de error está también estrechamente relacionada con la de problema de traducción y su importancia es fundamental no solo para la enseñanza sino también para la teoría de la traducción y para obtener más claridad en el proceso de análisis de la calidad de las traducciones. Sin embargo —afirma la autora— hasta ahora no hay una base sólida de estudios empíricos que sustenten la tipología de errores. De modo general se puede definir el error como una equivalencia inadecuada para el encargo de traducción (Hurtado 2011, 290). Para identificar los errores de traducción hay primero que tener en cuenta tres cuestiones principales:

1. La diferencia entre errores relacionados con el texto original y otros relacionados con el texto de llegada, aunque la autora reconoce que en algunos casos puede ser difícil distinguir si el error se da por un problema en la comprensión del texto original o por un deficiente en la redacción de la traducción.
2. La diferencia entre errores funcionales, de tipo pragmático que están relacionado con la transgresión de ciertos aspectos funcionales del proyecto de traducción, y errores absolutos, que se dan independientemente de la tarea traductora y son una transgresión injustificada de las reglas culturales, lingüísticas o del uso de la lengua de llegada.
3. La diferencia entre errores en el resultado de la traducción y los cometidos en el proceso de traducción.

Considerando estos aspectos la autora esboza la siguiente tipología: en primer lugar podemos encontrarnos con inadecuaciones que afectan a la comprensión del texto original: falso sentido, sin sentido, no dan el mismo sentido, adición, supresión, referencia extralingüística mal solucionada e inadecuación de variación lingüística (tono, dialectos, idiolecto, etc.). También pueden darse inadecuaciones que afectan a la expresión en la lengua de llegada: ortografía, puntuación, gramática, léxico, aspectos textuales (coherencia, progresión temática, elementos de referencia, conectores) y redacción (formulación defectuosa o poco clara, falta de riqueza expresiva, pleonasmos, etc.). Y por último podemos encontrarnos con inadecuaciones pragmáticas que no son acordes con la finalidad de la traducción (en relación con el

1. Traducción, traducción técnica, problemas de la traducción

tipo de encargo o destinatario), el método elegido, el género textual y sus convenciones, etc. (Hurtado 2011, 305–306).

Tipos de errores	Tipos de inadecuaciones
Aquellos relacionados con el texto original y/o con el texto de llegada	Afectan la comprensión del texto original
Funcionales o absolutos	Afectan a la expresión en la lengua de llegada
Errores se dan en el resultado o el en proceso de la traducción	Pragmáticas, que no son acordes con la finalidad de la traducción

Tabla 11: Clasificación de los errores e inadecuaciones de traducción según Hurtado

1.4.3 La clasificación de los errores de traducción según Schmitt

Los problemas en la traducción según este autor pueden tener diferentes causas y pueden darse entre otros, por *defectos* en el texto de partida. Si bien el autor no habla directamente de errores, podríamos afirmar, que la categorización que hace de los defectos se asemeja a una tipología de errores de traducción. Sin embargo la definición que Schmitt hace de los defectos no corresponde exactamente con la clasificación de errores de otros autores, porque el autor se centra en los problemas de traducción que surgen por defectos (o errores) en el texto original, es decir, errores que están en el texto antes que se inicie el proceso de traducción. Según Schmitt los defectos en el texto de partida generan un problema de traducción porque el traductor se encuentra en el dilema de ser fiel al original y ser consciente de que se está produciendo un texto erróneo en la traducción (Schmitt 1999, 60). Es por ello que incluso existen normas legales que eximen de responsabilidad al traductor en caso de que produzca una traducción errónea, debido a que el original estaba errado. Schmitt clasifica los principales tipos de defectos que se encuentran particularmente en los textos técnicos:

1) **Defectos formales**, tales como aquellos que se dan cuando un párrafo se repite dos veces, o cuando falta una parte o incluso una página en un texto, como un párrafo o capítulo en un manual de instrucciones, o cuando una referencia tiene un número de página errado.

2) Los **errores en las cifras o unidades de medida** indicadas constituyen también otra categoría frecuente de defectos en los textos originales. Errores en las unidades métricas

1. Traducción, traducción técnica, problemas de la traducción

se pueden detectar fácilmente. *Die 327 m lange Bundesbahn-Neubaustrecke Hannover-Würzburg (...) gilt als bedeutendste Bauvorhaben der Bahn seit Gründung der Bundesrepublik Deutschland.* Schmitt da este ejemplo para mostrar como en algunos casos con tan solo un mínimo de conocimiento de historia o geografía local puede detectarse errores de este tipo. Una autopista de 327 metros de longitud no puede calificarse como una de las obras de ingeniería más importantes en Alemania. En muchos casos el error no es tan evidente, y el reconocerlo como error depende generalmente del conocimiento que tenga el traductor en la materia.

3) También los **errores de ortotipografía o de imprenta** hacen parte de los errores o defectos en el texto de partida. Para reconocer este tipo de errores se requiere en algunos casos de un conocimiento en la terminología del tema en cuestión pero en otros casos solo de atención durante la traducción.

4) La **discrepancia entre el texto y las gráficas correspondientes** constituye también un error común. Puede suceder que el título de una imagen no sea correcto o corresponda a otra imagen; también podemos encontrar que en una imagen o en un modelo de un producto, los términos empleados no correspondan con la terminología empleada en la descripción adjunta del producto o el posicionamiento de la referencia sea incorrecto.

5) **Discrepancia entre texto y realidad.** Este tipo de errores se dan en algunos manuales de instrucciones entre otras cuando se describen partes de un aparato que tienen algún tipo de demarcación, que en realidad en el aparato es otra distinta (Schmitt 1999: 83–85).

6) **Errores de comprensibilidad.** En esta categoría no hay algo falso o incorrecto, se trata de ese tipo de errores o defectos en la formulación que hacen difícil la comprensión de un texto original. Es difícil en ciertos textos técnicos formular instrucciones o explicaciones, respetando las convenciones del tipo de texto, de tal manera que sean comprensibles para los receptores sin omitir información, pero tampoco sin dar más información de la necesaria. A este tipo de errores también pertenecen los errores en la **macroestructura**, es decir en el ordenamiento de las partes del texto en contexto global; también la **microestructura** es importante, el autor se refiere aquí al ordenamiento de las partes de un texto. Si la acción se da en una secuencia de pasos, estos deben estar claramente indicados como tales y tener cierta simi-

1. Traducción, traducción técnica, problemas de la traducción

litud, es decir, un paso no debe ser descrito más detalladamente que otro, a menos que sea estrictamente necesario o incluir varios pasos de acción en uno solo.

Errores que afectan la comprensibilidad de un texto también se dan **a nivel de las frases u oraciones**. Schmitt analiza el caso de la traducción la traducción al alemán del término inglés *actual*. Este término se utiliza en condiciones de garantía o de servicios de productos y su traducción puede tener diferentes implicaciones. También en castellano pueden tener tales palabras interpretaciones distintas. Veamos el siguiente ejemplo:

The right of (buyer's name) to claim damages for delay shall be conditional upon the submission of a written claim therefore within thirty days from the date of actual delivery of that Unit (Schmitt 1999: 89).

Según el autor no existe claridad con respecto con respecto a lo que quiere decir *date of actual delivery*, ya que no se puede definir con exactitud si se refiere a la fecha en que se entregó el artículo o si tiene que ver con la fecha en que se acordó entregar el artículo. De la interpretación de esta frase depende si el comprador tiene o no derecho de reclamar en caso de que no haya recibido el artículo.

A nivel morfológico también pueden presentarse errores especialmente si se utilizan polisemias. Como en el caso de *replace* en inglés que puede emplearse tanto para indicar que se debe cambiar algo, como para pedir que se vuelva a poner en un lugar determinado. Otra fuente de errores de comprensión se da con ciertas palabras o frases que tienen un significado específico si están acompañadas, pero cuya relación no es completamente clara, si no se utilizan guiones para señalar su relación. El ejemplo de Schmitt es *remove upper transmission case to block bolts*. Las palabras que tienen una relación lógica con *case-to-blocks*, pero si no están unidas por guiones se puede traducir la oración en un sentido diferente.

7) Schmitt clasifica como otra fuente de error el hecho de que ciertos autores no consiguen expresar lo que realmente querían comunicar. Es decir, **defectos de expresión**, que pueden darse porque el autor del texto no conoce la materia o porque aun siendo un experto en el tema no consigue transmitir la idea correctamente. Tal tipo de errores son muy frecuentes en textos técnicos, que son escritos por redactores, que saben escribir pero que no tienen

1. Traducción, traducción técnica, problemas de la traducción

un conocimiento técnico especializado en el tema, o porque son escritos por expertos cuyo manejo del idioma no es el más fluido.

8) Finalmente nos encontramos con **errores de contenido**. Para su explicación Schmitt da como ejemplo la oración en alemán *Das für VW gefertigte Mountainbike hat ein gelötete Aluminiumrahmen, Hohlkammerfelgen und in Ösen befestigten Nirosta-Speicher*. El autor se da cuenta de que existe un error en el contenido de la frase porque el adjetivo *gelöte*, que determina el complemento *Aluminunrahme* (marco de aluminio) indica en alemán que la parte en cuestión ha sido soldada, pero no es posible ya que aluminio no se puede soldar. El traductor debería utilizar el término correcto, en alemán el adjetivo indicado sería *geschweißt*, y traducir este y no el que el autor del texto original empleó (Schmitt 1999, 100–102).

El autor da otro ejemplo en el que encontramos un destornillador en el listado de requerimientos para la instalación de una tarjeta gráfica, al lado de las características del ordenador, la versión del sistema operativo y las capacidades del hardware requerido. Es claro que esta herramienta no se puede comparar con los otros puntos listados. El traductor deberá escoger otro título o eliminar el destornillador de la lista. Además de este tipo de errores evidentes se suelen pasar errores más sutiles. En la misma lista el autor descubre que las características técnicas que se le dan al ordenador, no son compatibles con el producto a instalar, es decir, con la tarjeta gráfica (Schmitt 1999, 100). Aquí cabe preguntarse si este tipo de errores o incongruencias se pueden considerar como errores que debe resolver el traductor, porque en realidad van más allá de su competencia.

Según el Schmitt el traductor debe conocer el *skopo* de la traducción para poder tomar la decisión correcta en caso de que se encuentre con defectos o errores en el texto original. Si el traductor no está seguro si se trata de un error o no, y no tiene la posibilidad de corroborarlo con el autor o redactor, puede al menos hacer un comentario o indicación al respecto. El traductor debe tener un conocimiento en la materia que le permita identificar los defectos o errores en el original y con ello no los transmita a la traducción. En la tabla siguiente resumiremos los errores o defectos que define el autor.

1. Traducción, traducción técnica, problemas de la traducción

Tipos de error	Ejemplo o caracterización
Defectos formales	Repetición u omisión de una parte del texto.
Errores en las cifras y/o unidades de medida	<i>Bei mittleren Wind soll das Schiff mühelos eine Geschwindigkeit von 10 km erreichen.</i> (Aquí se indica mal la medida de velocidad, se habla solo de km y no de km/h)
Errores de impresión y/u ortográfica que alteran el sentido	<i>F22 ist the force exerted by link 2 or link 3.</i> (correcto sería on)
Discrepancias entre el texto y las imágenes que le corresponden	El título de una imagen no es correcto o es el de otra imagen.
Discrepancias entre el texto y la realidad	El nombre o sigla en un dispositivo es diferente del que se encuentra en el dispositivo.
Errores o defectos en la comprensión	Se presentan p.ej. cuando hay errores de ordenamiento en partes o en todo el texto.
Defectos de expresión	El texto no es fácil de leer, es confuso.
Errores de contenido	Se emplea una traducción incorrecta.

Tabla 12: Clasificación de errores de Schmitt

La clasificación de errores de Schmitt es interesante, pero hay que recalcar que se trata más de una clasificación de defectos en los textos originales, que pueden generar problemas en la traducción, que en una clasificación de errores de traducción.

1.5 Problemas de traducción en la combinación lingüística alemán-castellano

Los problemas que hemos descrito en el apartado anterior no han sido formulados teniendo en cuenta una combinación lingüística específica. Es otras palabras, estos problemas son propios de la traducción independientemente de los idiomas que estén involucrados en el proceso concreto de traducción. En el apartado a continuación veremos cuáles de estos problemas se presentan más frecuentemente en la combinación lingüística de nuestra tesis.

1. Traducción, traducción técnica, problemas de la traducción

1.5.1 Problemas pragmáticos

En el capítulo anterior vimos como para autores como Nord existen cierto tipo de problemas que surgen cuando un texto, que funcionaba adecuadamente en la situación original, no funciona de la misma manera en la situación y/o en el idioma meta. Para resolver este tipo de problema hay que adaptar el texto a la situación o cultura meta sin cambiar por completo el original para que —a pesar de la distancia pragmática y cultural existente— el texto meta cumpla con la intención o la función con la que fue formulado originalmente.

Problemas en las referencias al autor, emisor o receptor de la traducción

En nuestra combinación lingüística pueden presentarse problemas según Nord al tener que traducir las referencias tanto al emisor o productor del texto, como al receptor. En algunos casos y dependiendo del tipo de traducción es necesario adaptarlas, por ejemplo si se trata de nombres que no tienen sentido en el contexto del idioma meta o si son nombres de instituciones que tienen una denominación diferente en el idioma meta. Este tipo de factores tienen que ver con las implicaciones y presuposiciones características tanto del texto original como del texto meta. En cuanto a las referencias que tienen que ver con el ‘hablante’ en el texto original, la autora enumera el registro pero también interjecciones o formas de saludo.

El traductor debe tener en cuenta, que aunque el texto original esté escrito de una manera particular, debe ser adaptado a la situación o cultura meta para que no suene extraño. Del mismo modo se deben tratar personales y las paráfrasis de éstos.

Problemas al interpretar y/o traducir la intención del autor

Nord (2009, 215) y Hurtado (2011, 288) se refieren a la intención comunicativa del autor como un aspecto que puede volverse problemático en la traducción. Para Nord es especialmente problemática la utilización del discurso indirecto, mediante el que el autor pretende influir de una determinada manera al receptor. Hay que tener en cuenta que la manera en que se puede influir al receptor depende de la situación y la cultura. Si un productor quiere vender un producto en un mercado alemán, deberá tener en cuenta qué tipo de información es más importante para el consumidor en dicho mercado. Si el mismo producto se va a vender en España se deberían tener en cuenta también los aspectos que son importantes en este mercado y cultura. Por ejemplo si en el texto original se resalta el hecho de que el producto tiene un

1. Traducción, traducción técnica, problemas de la traducción

manual de instrucciones en alemán, información que en el mercado alemán podría motivar el comprarlo, la intención original del autor de influir al cliente alemán no tendría sentido en el mercado español, si el manual no está traducido al castellano.

Problemas que se presentan por el medio

El medio en el sentido en que Nord lo emplea (2001a, 162) también puede constituirse en un problema pragmático en esta combinación lingüística. En el caso de que haya una limitación de espacio, como es el caso si estamos traduciendo un texto que se debe visualizar en el monitor de un teléfono. En alemán se emplean palabras compuestas, que no tienen un equivalente en castellano, y para cuya traducción en algunas ocasiones se tiene que crear un sintagma nominal y por ello en algunos casos la traducción tiene más palabras o letras. Es por eso que a veces nos encontramos en castellano con anglicismos o abreviaciones, que solo son utilizadas por necesidad, pero no porque sean la opción más correcta.

Nord también menciona como un problema pragmático de este tipo la utilización en un mismo texto de varios idiomas. Este problema concreto se da en nuestra combinación lingüística, ya que en alemán se suelen utilizar palabras o conceptos en inglés y algunas veces en francés, que son entendidos en la cultura origen pero que en castellano deben ser traducidos.

Problemas en partes específicas de los textos

Nord también habla de que pueden surgir problemas pragmáticos que tienen que ver con la función de partes del texto, como ejemplos o comparaciones, que deben ser adaptados a la situación o cultura meta, para que se pueda cumplir la función comunicativa del texto. Schmitt (1999) no se limita a partes del texto sino que indica que este tipo de problemas se da cuando la misma función comunicativa en la cultura base y meta se puede cumplir con tipos de textos diferentes. Si bien este tipo de problema pragmático no solo se da en nuestra combinación lingüística merece ser mencionado por su importancia y por la frecuencia en que se presenta.

1.5.2 Problemas culturales

A diferencia del inglés o el francés, el alemán y el castellano han tenido solo esporádicamente contacto (Navarro 1996, 181). Según García Yebra (1984) no hay casi interferencias entre estas lenguas porque entre ellas existe una distancia muy grande, más grande por ejem-

1. Traducción, traducción técnica, problemas de la traducción

plo que la existente entre el francés y el castellano y también porque el número de traducciones del alemán al castellano no es hasta el momento muy considerable, en comparación con otras lenguas como el francés o el inglés. Sin embargo es posible que la influencia de la lengua alemana aumente en los últimos tiempos, debido al papel dominante del país en la comunidad económica europea. Es así como nos encontramos a) con ‘nuevas’ palabras que se toman directamente del alemán, como *hansa* (asociación) o *Mittelsmann* (intermediario en una operación comercial), que se utilizan directamente sin ninguna adaptación en su escritura o pronunciación; b) también han aparecido préstamos, es decir palabras que se han naturalizado o adaptado al castellano como *camaralismo* que viene del alemán *Kammer* y con la que designa todo lo concerniente a la administración de ingresos o fastos de un príncipe; y c) calcos, que son la traducción directa, y que por ello aportan un concepto nuevo al idioma en que se utilizan, aquí podemos citar Banco federal (*Bundesbank*), milagro económico (*Wirtschaftswunder*) y unión aduanera (*Zollverein*) (Sánchez 2004).

Pero no solo en el ámbito económico encontramos influencias del alemán en el castellano, también encontramos germanismos como *Edelweis*, *Kaiser*, *Panzer*, adaptaciones como *Zepelin* (de *Zeppelin*) y calcos germanizados como el concepto filosófico de visión del mundo (*Weltanschauung*) el superego (*Überich*) de Sigmund Freud, el espíritu de la época (*Zeitgeist*) o el jardín infantil (*Kindergarten*) (Rodríguez Díaz 2011, 179).

A pesar de que no haya una gran cercanía entre estos dos idiomas, una clasificación general de los problemas culturales en la combinación lingüística alemán-castellano es bastante difícil de hacer, ya que existe una similitud cultural en ambas lenguas. Las diferencias concretas solo se pueden ver en un análisis detallado de textos y autores, y esta tarea va más allá de nuestros propósitos.

1.5.3 Problemas en el uso y aplicación de convenciones

Este tipo de problemas son definidos por Hurtado como textuales y por Nord como convencionales y algunas veces como culturales. A diferencia de los problemas pragmáticos, que se dan en cualquier combinación lingüística y cultural, esta clase de problemas tienen una relación directa con la combinación lingüística que se analice ya que como su nombre lo indica los problemas culturales o convencionales se originan en las diferentes normas de las cultu-

1. Traducción, traducción técnica, problemas de la traducción

ras y lenguas implicadas. Nord presenta un listado (no exhaustivo) de estos problemas de acuerdo a los siguientes complejos problemas: convenciones según el tipo de texto; convenciones generales de estilo; convenciones formales y otro tipo de convenciones bajo las que la autora agrupa las unidades de medida y las convenciones de traducción (Nord 2001a, 169).

Convenciones según el tipo de texto

Las convenciones de los tipos de texto son señales de reconocimiento, elementos desencadenantes de expectativas que orientan la comprensión del texto (Reiss y Vermeer 1996, 164–170). Bajo tipos de texto se agrupan los textos cuyas características son maneras de empleo de expresiones escritas u orales específicas de la situación que de mayor o menor manera se han institucionalizado o fijado en los actos comunicativos de una sociedad (Beck 1973, 73). Por ello mismo propone Nord que cada tarea de traducción debe estar precedida por el análisis de las convenciones del tipo de texto en el lenguaje a traducir.

Nord establece que hay determinados rasgos que caracterizan cierta clase de textos en alemán. Por ejemplo resalta, como textos como condiciones de término y negocios tienen tanto a nivel léxico como semántico diferentes características a textos del mismo tipo en castellano. A nivel léxico Nord observa que en textos de orden jurídico o legal en alemán existe una tendencia mayor a utilizar hiperónimos, compuestos nominales y recurrencias que en castellano. A nivel sintáctico se suelen utilizar en alemán atributos calificados con participios en lugar de aposiciones. También se prefieren las construcciones nominales mientras que en castellano se suelen utilizar más construcciones verbales; se emplean anáforas en lugar de oraciones relativas; se verbalizan los sustantivos en lugar de usar infinitivos, etc. (Nord 2001a, 171). En textos como condiciones de garantía lista la autora rasgos típicos de dichos textos en alemán, como son la estructura misma de los mismos, a nivel léxico la referencia en primera persona plural y la utilización de cierto tipo de expresiones típicas. También a nivel sintáctico se prefiere un registro formal y la utilización del estilo nominal.

En lo que se refiere a los textos técnicos tenemos que resaltar que en alemán está claramente definido cómo deben ser redactados y estructurados dichos documentos e incluso existe una reglamentación oficial para ello. Nos referimos a normas como DIN EN 61355-1 DIN V 8418 o 2006/42/EG en las que se describen que estructura y contenido deben tener la

1. Traducción, traducción técnica, problemas de la traducción

documentación de productos técnicos. Estas normas se basan en las directrices ISO 37 (Organización Internacional de Normalización) o del Instituto Americano ANSI, que define los estándares americanos. En España existe también un organismo privado encargado de la elaboración de normas UNE AENOR (Agencia Española de Normalización), sin embargo no existe ninguna norma técnica sobre manual de instrucciones (Gamero 2001, 96–98).

Nord se refiere también a las convenciones de otro tipo de textos como prospectos médicos, formularios, reseñas literarias, avisos inmobiliarios, recetas de cocina, eslogan, programas turísticos, títulos y partes del texto (Nord 2001a, 171–181); no entraremos aquí en detalle porque estos tipos de texto no hacen parte del corpus de nuestra tesis.

Convenciones de estilo

Como vimos anteriormente las convenciones generales de estilo tienen que ver con el léxico, la sintaxis y la constitución misma del texto y se vuelven relevantes para la traducción cuando hay diferencias en los idiomas implicados. Estas diferencias pueden ser cierto tipo de preferencias o la existencia de convenciones diferentes, también la frecuencia en que se emplean (Nord 2001a, 182). En nuestra combinación lingüística se da una diferencia en el uso de ciertos elementos estilísticos, como por ejemplo a nivel sintáctico se diferencian el alemán y el castellano en la utilización del estilo verbal y el nominal, lo que hace que si se traduce sin tener en cuenta el estilo, tengamos una traducción gramaticalmente correcta pero que no se ajusta a las normas de la cultura y el idioma. En alemán la aposición se encuentra antes o después del elemento gramatical a complementar. En castellano la aposición aparece generalmente después del elemento que complementa de lo contrario podría catalogarse como un error de estilo.

En castellano encontramos según Nord tres tipos de oraciones relativas: las necesarias o restrictivas, que no van precedidas de coma y especifican claramente la palabra a la que se refieren. Las explicativas, que se separan con comas y describen una propiedad o calidad adicional del objeto al que se refieren y las ilativas, que también se separan con comas, pero que no se refieren directamente a una palabra o elemento sino a toda la oración. Este tipo de oraciones también se dan en alemán pero no se pueden diferenciar formalmente. En alemán se utilizan menos oraciones relativas que en castellano y se emplean otro tipo de construcciones;

1. Traducción, traducción técnica, problemas de la traducción

por ejemplo cuando en castellano se utilizan oraciones relativas necesarias o restrictivas, en alemán suele emplearse atributos o complementos preposicionales (Nord 2001a, 185).

Castellano	Alemán
Esta tarea que nos repugna	<i>Diese wiederwärtige Arbeit</i>
Las ventajas que se derivan de ser viviendas de Protección Oficial	<i>Die Vorteile des öffentliche geförderten Wohnungsbau</i>

Tabla 13: Ejemplo del uso de oraciones relativas en castellano y alternativas en alemán

En cuanto a las convenciones formales en alemán no suelen marcarse tipográficamente los extranjerismos, mientras que en castellano sí. Como en alemán se escriben todos los sustantivos con mayúscula, no se puede utilizar la mayúscula como marcación especial o exclusiva de los nombres propios. Además en alemán existen ciertas convenciones que no encontramos en castellano:

Alemán		Castellano	
Las cantidades se escriben con cifras o fracciones:	<i>1/2 Liter Milch und 2 Essl. Zucker</i>	Se utilizan cifras pero también suelen transcribirse.	un litro, medio litro
Los números ordinales se colocan antes del sustantivo:	<i>2. Phase des Projekts</i> ”.	Se colocan después del sustantivo	'fase 2 del proyecto'

Tabla 14: Convenciones formales en alemán (Nord 2001a, 189)

Nord también se refiere a otro tipo de convenciones formales y también a convenciones literarias que no especificaremos porque no son relevantes para nuestra tesis.

1.5.4 Problemas lingüísticos

Como vimos anteriormente según autores como Nord y Hurtado existen problemas de traducción que están relacionados con el código lingüístico y que se dan tanto en el plano léxico, como en el morfológico. Siguiendo esta distinción en los planos, listaremos a continua-

1. Traducción, traducción técnica, problemas de la traducción

ción los problemas de orden lingüístico más comunes que encontramos al traducir del alemán al castellano. No podemos excluir que se nos escapen algunos problemas en este listado.

1.5.4.1 Problemas léxico-semánticos

En el orden léxico-semántico la ambigüedad y la polisemia ocasionan problemas y errores en la traducción. También el uso de colocaciones o metáforas son problemáticos en las traducciones y pueden ser causa de errores, que muchas veces pueden afectar la traducción completa de un texto. Aunque estos no son problemas exclusivos de nuestra combinación lingüística, mostraremos a continuación porque en la traducción del alemán al castellano son tratados frecuentemente.

La ambigüedad

En el ejemplo siguiente (Hans Uszkoreit 1998) podemos ver que dimensiones puede alcanzar el problema de la ambigüedad en nuestra combinación lingüística:

*Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.*²

Elemento	Posibles interpretaciones	Número
<i>Früher</i>	puede ser un adverbio o el comparativo del adjetivo <i>früh</i>	2
<i>stellten</i>	La forma verbal puede estar tiempo pretérito o en modo subjuntivo	2
<i>die Frauen</i>	el sintagma nominal pueden ser objeto o sujeto en la oración	2
<i>am Wochenende</i>	puede modificar <i>die Insel</i> , <i>die Frauen</i> o al verbo (<i>her</i>) <i>stellen</i>	3
<i>mit Blumenmotiven</i>	puede referirse a <i>Kopftücher</i> , ser un instrumento de fabricación o un complemento	3
<i>her</i>	puede ser una partícula que indica una dirección o hacer parte del verbo <i>stellen</i>	2

² Antes, las mujeres de la isla fabricaban los fines de semana pañuelos para la cabeza con motivos florales, que sus esposos vendían los lunes siguientes en el mercado en el centro de la isla principal. (La traducción es nuestra)

1. Traducción, traducción técnica, problemas de la traducción

Elemento	Posibles interpretaciones	Número
<i>die ihre Männer ...verkauften.</i>	La oración relativa puede modificar cada una de los sintagmas nominales en plural	4
<i>Die / ihre Männer</i>	pueden ser sujetos de la oración relativa	2
<i>ihre</i>	el pronombre posesivo puede referirse a cada uno de los sintagmas nominales	4
<i>Montagen</i>	pueden interpretarse de dos maneras distintas: plural del día de la semana o nominalización del verbo <i>montieren</i>	2
<i>der Hauptinsel</i>	puede pertenecer al sintagma nominal anterior (genitivo) o referirse a las compradoras (dativo)	2
	los tres sintagmas preposicionales de la oración relativa pueden tener 7 combinaciones con los sintagmas nominales anteriores o con el verbo	7
<i>verkauften</i>	el verbo puede estar en pretérito o conjuntivo	2
	$2 \times 2 \times 2 \times 3 \times 3 \times 2 \times 4 \times 2 \times 4 \times 2 \times 2 \times 7 \times 20$	= 258.048

Tabla 15: Ejemplo de ambigüedad de Uszkoreit

Si tenemos en cuenta estas posibles combinaciones la oración podría tener 258.048 interpretaciones. El que se pueda interpretar la oración de tantas maneras diferentes se debe a que existen ambigüedades de diferentes tipos, no solo ambigüedad léxica, sino también sintáctica.

La ambigüedad léxica puede darse porque unidades léxicas que pueden tener al menos parcialmente la misma forma pueden tener significados diferentes, hablamos aquí de los homónimos; también la polisemia, palabras que tienen la misma forma, pero pueden tener parcialmente diferentes significados generan una ambigüedad léxica. Además en ciertas figuras como metáforas o colocaciones puede darse ambigüedad de carácter léxico.

1. Traducción, traducción técnica, problemas de la traducción

Homónimos

Nos encontramos con homónimos totales, como en el caso del sustantivo alemán *Weiche* (Löbner 2003, 58-59), en los que todas las características del lexema como género o número coinciden, pero tienen un significado diferente.

Sustantivo	Significado
<i>Weiche</i>	Cambiavía del tren
	Ruta alternativa
<i>Weiche</i>	Parte del cuerpo
	Consistencia

Tabla 16: Ejemplos de homonimia total en alemán

También tenemos casos de homonimia parcial donde algunas de las características, por ejemplo el género o la formación del plural pueden ser diferentes y no generar ambigüedad:

Sustantivo (singular)	Sustantivo (plural)	Significado
<i>Bank</i>	<i>Bänke</i>	Mueble
<i>Bank</i>	<i>Banken</i>	Instituto financiero

Tabla 17: Ejemplo de homónimos parciales en alemán (Löbner 2003, 58-59)

Algunos casos de homonimia parcial los encontramos en los homógrafos, palabras que se escriben igual pero que tienen un significado diferente y por ello pueden ocasionar problemas de traducción. En este caso hablamos también de **ambigüedad ortográfica**. Este es el caso del verbo alemán *übersetzen* al que se le pueden atribuir dos significados distintos: por un lado puede interpretarse como traducir de un idioma a otro, pero también como pasar/llevar algo o alguien de un lado u orilla a otro³. Esta ambigüedad en particular se originó porque a partir del siglo XVII se utilizó la palabra del alto alemán medio (alemán antiguo *ubarsezzen*) como correspondiente a las palabras latinas *traducere* o *transferre*. Como en este caso, la ambigüedad puede resolverse en el contexto, sin embargo no se puede excluir que

³Este verbo tiene un tercer significado técnico: transformar la velocidad líneas de rotación o fuerza de un motor.

1. Traducción, traducción técnica, problemas de la traducción

puedan presentarse problemas de traducción. Palabras homófonas también pueden considerarse como casos de homonimia parcial, pero no los tendremos en cuenta ahora, porque solo nos ocuparemos de textos escritos.

No solo los morfemas o palabras pueden generar ambigüedad, también composiciones complejas pueden ser ambiguas dependiendo de sus estructuras lógico-semánticas:

Original en alemán	Posible interpretación
<i>Jeder beneidet jemand</i>	Todas las personas envidian a una persona o cualquiera es envidiado por otro

Tabla 18: Ambigüedad semántica en una oración (Ernst 2003: 79)

Especial del alemán son los sustantivos compuestos, en los que el primero a veces está en plural y el segundo es un verbo nominalizado. Aquí no queda claro si el verbo es activo o pasivo:

Original en alemán	Posible interpretación
<i>Das Bauernopfer</i> (Reischer 2002, 268)	a) El campesino es sacrificado
	b) El campesino hace un sacrificio
<i>Der Fenstersturz</i>	a) Se cae una ventana
	b) Alguien o algo cae de una ventana
<i>Der Bocksprung</i>	a) Salta alguien encima del potro
	b) el potro salta

Tabla 19: Ambigüedad de nombres compuestos en alemán

Polisemia

Otro caso de ambigüedad léxica lo constituye la polisemia. A diferencia de los homónimos, en los que diferentes lexemas tienen una forma idéntica, en la polisemia un lexema tiene diferentes variantes de significado (Löbner 2003, 62). Es decir a la polisemia representa la posibilidad de entender de diferentes maneras un morfema, una palabra, una secuencia, una oración, etc., independientemente de si esta múltiple interpretación está relacionada o no con

1. Traducción, traducción técnica, problemas de la traducción

cierta descripción gramatical, por ejemplo si solo se presenta aislada o en un determinado contexto.

La polisemia no es un fenómeno exclusivo de los idiomas de nuestra combinación lingüística, pero es innegable que puede ocasionar problemas en la traducción del alemán al castellano. Mientras que los homónimos pueden tener diferente significado y en muchos casos un origen etimológico diferente, en los casos de polisemia, generalmente encontramos una raíz común, a partir de la que se hace la diferencia en el significado. Hablamos, por ejemplo del “caballo” para designar tanto el animal, como la ficha del juego de ajedrez.

También se puede llegar hasta el punto de que el significado en una frase sea vago y que un mismo significante se puedan interpretar de maneras incluso opuestas o contradictorias. Por ejemplo, la oración en alemán: *Das Auto wird das Hindernis umfahren*. En este caso hay una ambigüedad de carácter léxico en el significante *Umfahren*. La oración se puede interpretar y traducir en dos sentidos distintos y contradictorios, ya que el verbo *umfahren* puede interpretarse aquí de dos maneras distintas:

Original en alemán	Significado	Ejemplo
<i>umfahren</i>	evitar algo	El coche se desvía y evita el obstáculo
	atropellar algo o alguien	El coche no se desvía y choca contra el obstáculo

Tabla 20: Ejemplo de polisemia en alemán

En la traducción del alemán al castellano también nos confrontamos con los problemas que pueden ocasionar los diferentes tipos de polisemia. Entre otras porque no solo nos vemos confrontados con la ambigüedad del texto base, que tenemos que resolver, sino porque en muchos casos, la polisemia no existe en alemán y se debe adaptar el sentido.

Mientras que la polisemia regular, puede interpretarse y traducirse correctamente, la traducción de los casos de polisemia que se presentan por ampliación de significado o especialización deben ser adaptados a la cultura meta.

1. Traducción, traducción técnica, problemas de la traducción

Tipo de polisemia	Original en alemán	Traducción	Explicación
Normal (Bierwisch 1983)	<i>Die Schule wird neu eröffnet</i>	La escuela se inaugurará hoy	Escuela = institución
	<i>Die Schule hat ein rotes Dach</i>	La escuela tiene un techo rojo	Escuela = edificación
	<i>Die Schule endet heute früher</i>	La escuela acaba hoy antes	Escuela = concepto que representa un proceso
Por ampliación del significado:	<i>Peter ist eine Bohnenstange</i>	Pedro es un rodrigón	<i>Bohnenstange</i> = rodrigón
		Pedro es alto y flaco	rodrigón = largo = flaco
Por especialización	<i>Surfen auf Hawaii</i>	hacer surf en Hawaii	
	<i>Im Internet surfen</i>	navegar en Internet	
	<i>Die Maus ist tot</i>	Se ha muerto el ratón	ratón = animal
		Se ha muerto el ratón	ratón = del ordenador

Tabla 21: Tipos de polisemia Metáforas y metonimias

El desplazamiento intencional del sentido de una expresión se llama metonimia y el cambio asociado en el sentido desplazamiento metonímico. Las metonimias y las metáforas son utilizadas también en alemán y pueden constituirse en problemáticas cuando el sentido no se puede reproducir en castellano. Mientras que las metonimias tienen un carácter más o menos universal, muchas metáforas están arraigadas en la cultura y no pueden traducirse simplemente, ya que la metáfora no suele crear conceptos nuevos sino transcribir de manera subjetiva de cómo éstos son experimentados (Mellado Blanco 1999).

Por ejemplo mientras que en alemán *von Gott verlassen sein* (abandonado por Dios) significa haberse vuelto loco, la expresión similar en castellano “dejado de la mano de Dios” estaría designando a una persona que ha sido abandonada o desatendida (Mellado Blanco 2009, 57). Interesante es también el uso frecuente en alemán de metáforas con referencia a animales, cuyo sentido es generalmente negativo:

1. Traducción, traducción técnica, problemas de la traducción

Expresión en alemán	Traducción	Significado
<i>Rabeltern</i>	Padres de cuervos	padres desnaturalizados
<i>Schweinehund</i>	Perro-cerdo	canalla
<i>Warteschlange</i>	Serpiente de espera	fila larga
<i>alte Eule</i>	lechuza vieja	mujer de edad y además molesta
<i>dumme Zicke</i>	Cabra boba	mujer boba o molesta

Tabla 22: Metáforas en alemán

Colocaciones y frases idiomáticas

Colocaciones y frases idiomáticas deben ser adaptadas al lenguaje meta. Esta tarea debe ser realizada en la traducción de cualquier idioma a otro y no vemos problemas particulares que solo se den en la combinación alemán-castellano, que no hayan sido contemplados en las otras categorías de problemas.

Vaguedad léxica

La vaguedad léxica, es decir el hecho de que por ejemplo una palabra pueda tener un significado que no es preciso dentro de una frase o enunciado puede ocasionar también problemas en la traducción. La denotación de un concepto tiene a menudo fronteras fluidas. Mientras que la ambigüedad puede darse si una palabra tiene varios usos o pos extensión de su significado, la vaguedad puede darse por falta de precisión, por el uso relativo de una palabra —por ejemplo cuando se utiliza un adjetivo calificativo en un juicio de valor como cuando se habla de ‘grande’ o ‘pequeño’ ya que la vaguedad se presenta en conceptos cuyas características tienen un valor que puede variar — o porque el significado mismo va cambiando en el tiempo o en el uso mismo de este. (Löbner 2003, 62).

1.5.4.2 Problemas de orden morfo-sintáctico y gramatical

Gracias a las declinaciones, el orden sintáctico en alemán es más reconocible que en muchos otros idiomas, incluyendo el castellano, lo que puede hacer aparentemente más fácil la traducción. En las secciones siguientes trataremos de mostrar qué tipo de problemas de traducción se presentan en nuestra combinación lingüística debido a las diferencias morfológicas, sintácticas y gramaticales de estos idiomas.

1. Traducción, traducción técnica, problemas de la traducción

Ambigüedad sintáctica

También encontramos otro tipo de ambigüedades que se originan por una ambigüedad morfológica ya que algunas secuencias de morfemas o palabras pueden interpretarse como palabras diferentes. Sin el contexto de una frase no es claro el significado. Este tipo de ambigüedad lo encontramos a menudo en alemán especialmente en palabras que tienen su origen en otro idioma y que han sido adaptadas al alemán con el paso de los años, conservando algunas veces el sentido original y otras el idiomático (Hutchins y Somers 1992, 85):

Sustantivo	Componentes	Significado
<i>Alleinvernehmen</i>	<i>Allein + Vernehmen</i>	Percepción aislada
	<i>All + Einvernehmen</i>	Acuerdo global
<i>Beileid</i>	<i>Bei + Leid</i>	Condolencia
	<i>Beil + Eid</i>	Juramento
<i>Kulturgeschichte</i>	<i>Kultur+ Geschichte</i>	Historia de la cultura
	<i>Kult + Urgeschichte</i>	Prehistoria del culto
<i>Uranbrenner</i>	<i>Ur+Anbrenner</i>	Quemador primitivo ⁴
	<i>Uran+brenner</i>	Pila de uranio
<i>Wachtraum</i>	<i>Wach+Traum</i>	Fantasía diaria (<i>day dream</i>)
	<i>Wacht+Raum</i>	Sala de guardia

Tabla 23: Ejemplos de ambigüedad en sustantivos compuestos en alemán de Hutchins y Somers

También otros autores nos dan ejemplos de este tipo de ambigüedad en alemán:

Sustantivo		Componentes	Significado
<i>Staubecken</i> (Görz, y <i>Schneeberger 2003, 58</i>)	<i>Stau-b</i>	<i>Stau + Becken</i>	embalse
	<i>Staub-e</i>	<i>Staub + Ecken</i>	esquinas llenas de polvo

⁴ Este ejemplo de los autores no aparece en el diccionario alemán (DUDEN)

1. Traducción, traducción técnica, problemas de la traducción

Sustantivo		Componentes	Significado
<i>Hauptpostsekretär</i> (<i>H. Uszkoreit 2002</i>)	-sekretär	<i>Haupt+post+sekretär</i>	asistente
	-sekretär	<i>Haupt+post+sekretär</i>	mueble (escritorio)

Tabla 24: Ambigüedad léxica debido a la estructura morfológica en alemán

Adicionalmente nos encontramos con casos ambigüedad no solo en morfemas sino también en oraciones, podemos hablar aquí de ambigüedad sintáctica ya que una oración o frase puede analizarse de dos o más maneras diferentes. Por ejemplo:

Original en alemán	Posible interpretación
<i>Briefträger beißen Hunde selten</i>	No se puede determinar claramente el sujeto de la acción:
	a) Los perros (<i>Hunden</i>) rara vez muerden a los carteros
	b) Los carteros (<i>Briefträger</i>) muerden rara vez a los perros
<i>Ich sah der Mann mit dem Fernglas</i>	No se puede determinar qué tipo de complemento es:
	a) Vi al hombre con (ayuda de) los prismáticos
	b) Vi al hombre que tenía en sus manos unos prismáticos

Tabla 25: Ejemplos de ambigüedad sintáctica

Este tipo de oraciones que tienen una ambigüedad sintáctica global, pues a pesar del análisis no se puede resolver la ambigüedad y para hacerlo se necesita de un contexto más amplio. Pero también nos encontramos con otro tipo de ambigüedad las oraciones “Garden Path” (Meng y Bader 2000). En este tipo de oraciones se da una ambigüedad local, que se puede resolver a través de otra palabra que aparece posteriormente en la oración en cuestión. Este tipo de oraciones se llaman así haciendo referencia a la expresión *to lead someone down the garden-path* y por ello pueden presentarse problemas en la asignación de una estructura sintáctica o *parsing*.

1. Traducción, traducción técnica, problemas de la traducción

Original en alemán	Posible interpretación
<i>Ich glaube, dass Max zugunsten von Tim etwas unternommen wurde.</i>	La ambigüedad comienza con el nombre propio Max:
	a) Max es el sujeto: Creo que Max hizo algo a favor de Tim
	b) Max es el objeto de la preposición <i>zugunste</i> : Creo que a Max le sucedió algo que favoreció a Tim
	Correcta es la opción b.) aunque una primera interpretación (errónea) favorecería la opción a)

Tabla 26: Ejemplo de una oración ambigua Garden Path (Bader 2004, 8-9)

En alemán se usan muchas construcciones con la forma nominalización + referencia + sustantivo que generan también fuera de contexto una ambigüedad sintáctica:

Original en alemán	Posible interpretación
<i>Das Fühlen der Hand... (Fränkel 2008, 15)</i>	a) El sentir la mano (la sensación que tiene una persona por su mano).
	b) La sensación de la mano (la sensación que percibe la mano: frío, calor, dolor).

Tabla 27: Ejemplos de ambigüedad nominalización + referencia + sustantivo

Construcciones en las que se emplea un sustantivo en plural y el verbo en infinitivo pueden generar ambigüedad:

Original en alemán	Posible interpretación
<i>Wilde Tiere jagen...</i>	a) los animales salvajes cazan. b) los animales salvajes son cazados.
<i>Plakaten kleben</i>	a) los pósteres deben ser pegados. b) los pósteres se pegaron, p.ej. unos con otros.
<i>Flugzeuge fliegen / Schiffe segeln / Autos fahren</i>	Aviones vuelan o volar aviones / Barcos navegan o navegar en barcos / coches andan o andar en coches.

Tabla 28: Construcciones ambiguas con sustantivo en plural y verbo en infinitivo

1. Traducción, traducción técnica, problemas de la traducción

También algunos caracteres complejos pueden tener varios significados. La diferencia en el significado depende de la ambigüedad de los signos léxicos individuales o de la ambigüedad sintáctica de la frase correspondiente:

Original en alemán	Posibles interpretaciones
<i>Jeder Mann liebt eine Frau.</i>	a) Cada hombre ama por lo menos a una mujer.
	b) Cada hombre ama solo a una mujer.
	c) Todos los hombres aman a una y a la misma mujer.
	d) Solo hay una mujer y todos los hombres la aman.
<i>Paul hat Maria nur zu seinem Fest eingeladen.</i>	a) Paul solo invito a María a la fiesta (y no a cine).
	b) Paul solo invito a María a la fiesta (él no la beso).
<i>Zeichnen Sie ein Bild von sich, nackt.</i>	a) Pinte un autorretrato desnudo.
	b) Pinte desnudo un autorretrato.
<i>Peter fuhr seinen Freund sturzbetrunken nach Hause.</i>	a) Peter llevo a su amigo, que estaba ebrio, a la casa.
	b) Peter, estando el mismo ebrio, llevo a su amigo a la casa.

Tabla 29: Ambigüedad estructural

Además presenta ambigüedad si no se puede concluir a partir del contexto a qué se refiere una palabra o frase. Este caso de ambigüedad es un tanto difícil de clasificar, pues la ambigüedad también depende de la estructura lingüística del sintagma. Este tipo de ambigüedad se da mucho en los textos en alemán ya que la flexibilidad sintáctica permite colocar los complementos con mayor libertad que en otros idiomas y tal libertad permite interpretar la frase de diferentes maneras.

Diferencias en la conjugación

En castellano la conjugación de los verbos es más compleja, no solo debido a la existencia de los tres tipos, sino por la proliferación de verbos irregulares. En alemán solo hay dos tipos de verbos, los débiles cuya raíz no cambia y los fuertes, cuya raíz se modifica. Los verbos siempre tienen la terminación que es sustituida en la conjugación por el sufijo o morfema gramatical.

1. Traducción, traducción técnica, problemas de la traducción

En alemán y castellano existen diferentes modos. Por ejemplo en castellano no tenemos el modo *Konjunktiv I* o discurso indirecto, que se utiliza especialmente por escrito para hacer énfasis en la afirmación que hizo un tercero. Al no haber un equivalente exacto en castellano puede presentarse problemas en la traducción, especialmente en el caso del discurso indirecto.

Discurso indirecto	<i>Mein Bekannter sagt, er habe geheiratet</i> (Mi conocido dijo, que se había casado).
Preguntas indirectas	<i>Sie wollten wissen, ob ich den Film gesehen habe.</i> (Ellos querían saber si había visto la película).
Deseo indirecto	<i>Der Gläubiger stellt beim Gerichtsvollzieher den Antrag, dass die Zwangsvollstreckung betrieben werde.</i> (El acreedor pide al agente judicial que se ejecute el apremio judicial).

Tabla 30: *Konjunktiv I* en alemán

Al *Konjunktiv II* le corresponden en castellano el subjuntivo y el condicional. En oraciones irreales se utilizan en castellano el subjuntivo en la oración principal y el condicional en la oración subordinada. Si se trata de deseos inalcanzables u oraciones comparativas reales se usa en castellano el imperfecto de subjuntivo o el pluscuamperfecto de subjuntivo (perfecto condicional). Esta distinción sino se realiza correctamente puede generar problemas de traducción.

Oraciones irreales	<i>Hätte ich das gewusst, wäre ich nicht gekommen.</i> Si lo hubiese/hubiera sabido no habría venido.
Deseos inalcanzables	<i>Wäre er wenigstens höflich gewesen!</i> ¡Si por lo menos hubiera sido amable!
Comparaciones reales	<i>Er kann pfeifen, als ob er ein Vogel wäre.</i> Él sabe silbar como si fuera un pájaro.

Tabla 31: *Konjunktiv II*

1. Traducción, traducción técnica, problemas de la traducción

A diferencia del alemán en castellano distinguimos con ayuda del verbo ser y estar las características inherentes al sujeto y aquellas que no lo son; esta característica del castellano ocasiona problemas de traducción ya que en ese idioma no se da esta diferenciación:

Original en alemán	Traducción al castellano
<i>Sie ist hübsch</i>	Ella es bonita
<i>Sie ist krank</i>	Ella está enferma

Tabla 32: Verbo ser / estar

El gerundio es una forma verbal que no existe en alemán. En combinación con el presente de indicativo del verbo estar, el gerundio puede asumir diferentes funciones: describir actividades que se están realizando en el momento; un desarrollo gradual “se está acostumbrando”; describir una acción o situación temporal de carácter temporal “estamos viviendo en Madrid”.

En contraste con el gerundio sencillo, el gerundio compuesto se refiere a hechos que ya han sido culminados. Esta forma se construye con el gerundio del verbo haber y el participio del verbo cuya acción se completa (el fiscal, habiendo valorado previamente el presupuesto, afirmó que...).

El alemán es una lengua en la que los verbos finitos de una oración enunciativa deben aparecer en la segunda posición en la frase. La primera posición puede ser ocupada por cualquier otro elemento de la oración. En los imperativos e interrogaciones el verbo se colocará en la primera posición. En todas las subordinaciones el verbo conjugado se coloca al final de la frase.

Las formas verbales compuestas como el perfecto, también demuestran otro orden sintáctico en ambos idiomas. Mientras en castellano el verbo auxiliar “haber” nunca se separa del participio, en el alemán sí se separa en las frases principales. En ambos idiomas se puede invertir la posición del sujeto (posición 3) para destacar un objeto o complemento:

1. Traducción, traducción técnica, problemas de la traducción

Original en alemán	Traducción
<i>Ich habe heute Fußball gespielt</i>	Yo he jugado a fútbol hoy
<i>Heute habe ich Fußball gespielt</i>	Hoy he jugado a fútbol

Tabla 33: Formas verbales compuestas y posición del verbo

El orden de los objetos y complementos

Como podemos ver en el ejemplo siguiente en alemán se ordenan los complementos de una manera diferente que en castellano:

<i>Der Lehrer</i>	<i>gab</i>	<i>meiner Mutter</i>	<i>einen Brief</i>
Sintagma nominal	Verbo	Complemento indirecto	Complemento directo
El maestro	entregó	una carta	a mi madre
Sintagma nominal	Verbo	Complemento directo	Complemento indirecto

Tabla 34: Ejemplo de la diferencia en el ordenamiento gramatical

Como vemos en castellano el objeto directo va delante del indirecto. En la frase principal en alemán el verbo siempre está en segundo lugar y le siguen el objeto indirecto (generalmente en dativo) y luego el complemento u objeto directo en acusativo. Si uno de los dos complementos es remplazado por un pronombre, el pronombre (sin importar si remplace al objeto directo o al indirecto) irá antes del complemento explícito. Si ambos complementos son remplazados por pronombres el orden se cambia:

<i>Der Lehrer</i>	<i>gab</i>	<i>ihn</i>	<i>ihr</i>
Sintagma nominal	Verbo	Complemento directo	Complemento indirecto
El maestro	la	entregó	a ella
Sintagma nominal	Complemento directo	Verbo	Complemento indirecto

Tabla 35: Ejemplo de la diferencia en el ordenamiento gramatical

Los complementos circunstanciales se colocan en alemán después del verbo conjugado (en oraciones simples) o entre el verbo conjugado y el infinito en el siguiente orden: tempora-

1. Traducción, traducción técnica, problemas de la traducción

les, causales, modales y por último los locales. Cabe resaltar que el orden se puede variar depende de la importancia que se le quiera dar a uno de los complementos y que es diferente en ambos idiomas.

Sujeto	Verbo	Temporal	Causal	Modal	Lugar	Directo	Verbo
<i>Herr Meier</i>	<i>hat</i>	<i>gestern</i>	<i>aus Liebe</i>	<i>im Geheim</i>	<i>in München</i>	<i>Blumen</i>	<i>gekauft</i>

Temporal	Sujeto	Verbo	Directo	Modal	Causal	Lugar	
<i>Ayer</i>	<i>el Sr.Meier</i>	<i>compró</i>	<i>flores</i>	<i>en secreto</i>	<i>por amor</i>	<i>en Munich</i>	

Tabla 36: Diferencias en la colocación de los complementos en alemán y castellano

La negación

En alemán existen diferentes partículas para negar una afirmación: principalmente *nein* o *nicht*. *Nein* se utiliza para responder negativamente a una pregunta, con *nicht* se niega un verbo, un adjetivo o un sustantivo. También se utiliza el pronombre *kein*, que puede ir acompañando al sujeto correspondiente (se declina como el artículo determinado= o no ser tener ninguna función atributiva (en el caso la declinación es fuerte.

Mientras que en castellano la posición de la partícula negativa ‘no’ es relativamente flexible, en alemán se debe colocar exactamente al frente de lo que se está negando.

Posición	Ejemplo y traducción
Frente al complemento	<i>Ich möchte nicht jeden Tag Reis essen</i>
	No quiero comer todos los días pasta (lo negado es todos los días)
Frente al verbo	<i>Ich möchte nicht essen</i>
	No quiero comer (lo negado es la acción misma de comer)

Tabla 37: Diferencias en la colocación de la negación en alemán y castellano

La traducción de la negación es problemática, ya que si la estructura sintáctica es compleja se puede interpretar erróneamente lo que se desea negar. Interesante también como

1. Traducción, traducción técnica, problemas de la traducción

diferencia es que en alemán no se utiliza la doble negación del castellano (no existen construcciones como “no tengo ningún”).

Las composiciones en alemán

En el idioma alemán encontramos muchas palabras compuestas. Existen sustantivos, verbos e incluso adjetivos compuestos. Este tipo de composiciones por regla general no tienen ninguna palabra correspondiente en castellano. Para su traducción se necesita más de una palabra e incluso en muchos casos hasta de una frase completa. A continuación daremos algunos ejemplos que dejan ver claramente como la traducción al castellano de tales composiciones puede ser problemática.

Característico del alemán es el empleo de los nombres compuestos, en los que el primero a veces está en el plural y el segundo es un verbo nominalizado y donde no está claro si el verbo es el uso activo o pasivo, si el contexto en que se encuentra el nombre no es claro o falta, puede optarse erróneamente por la traducción incorrecta:

Original en alemán	Posibles significaciones
<i>Bauernopfer</i>	a) el agricultor se sacrifica
	b) sacrificio de peón (jugada de ajedrez)
<i>Fenstersturz</i>	a) el dintel
	b) la defenestración (por ejemplo, la defenestración de Praga)

Tabla 38: Ejemplo de nombres compuestos en alemán con diferentes significados

En alemán existen muchos verbos compuestos por una raíz y un prefijo, cuyo significado puede variar dependiendo del prefijo que lleve. Este es el caso por ejemplo de los verbos siguientes: *herstellen* / *einstellen* / *anstellen* / *abstellen*.

Verbo en alemán	Significado	Ejemplo
<i>herstellen</i>	establecer	<i>Stellen Sie eine WLAN-Verbindung zwischen dem Repeater und Ihrer WLAN-Basisstation her.</i>
		<i>Establezca</i> una conexión inalámbrica entre el repetidor y su estación inalámbrica.

1. Traducción, traducción técnica, problemas de la traducción

Verbo en alemán	Significado	Ejemplo
	fabricar	Syrien kann keine Chemiewaffen mehr herstellen Siria no puede fabricar más armas químicas
<i>einstellen</i>	adaptarse	<i>Bürger müssen sich auf Behinderungen einstellen</i> Los ciudadanos tienen que adaptarse a las incomodidades
	contratar a alguien	<i>Bill Yosses wurde noch von Obamas Vorgängerin Laura Bush (67) eingestellt.</i> Bill Yosses ya había sido contratada por Laura Bush (67), la predecesora de Obama.

Tabla 39: Verbos compuestos en alemán con diferente significado

Otro tipo de problemas

El castellano como lengua descendiente del latín vulgar perdió las declinaciones y solo los pronombres personales manifiestan variaciones formales de caso. Por el contrario en alemán existen los casos que son una marca de función que establece la relación gramatical entre dos elementos. Dicha marca puede manifestarse morfológicamente (a través de variaciones en la flexión de los nombres, adjetivos, pronombres, etc.) o puede ser abstracta, en el sentido de no ser visible morfológicamente.

Mientras que en castellano la relación de los objetos (directo o indirecto) es especificada por preposiciones o por la colocación en la frase, en alemán se utilizan las declinaciones. Esto se ve más claramente en el uso del genitivo en alemán. En castellano empleamos generalmente la preposición “de” en alemán el genitivo. Algo parecido sucede con el dativo, que podemos remplazar en castellano por la preposición “a”.

Original en alemán (genitivo)	<i>Der Mann schenckt das Auto dem Kind seines Freundes</i>
Traducción al castellano (preposición)	<i>El hombre regala el auto al hijo de su amigo</i>

Tabla 40: Uso de las declinaciones en alemán

También el artículo indefinido cambia en alemán dependiendo del caso mientras que en castellano se utilizan las preposiciones para especificar el caso.

1. Traducción, traducción técnica, problemas de la traducción

<i>Das Buch einer Frau</i>	El libro de una mujer
<i>Ich schenke einigen Frauen Geld.</i>	Regalo dinero a unas mujeres.

Tabla 41: Declinación del artículo indefinido en alemán

La formación del plural en alemán es bastante compleja. Para ello se utilizan diferentes sufijos -e, -n, -er o -s. El sufijo más usado es -e. La mayoría de las palabras que forman el plural con el sufijo -s son provenientes del inglés. En algunos casos no solo se añade un sufijo sino que además a la vocal de la raíz se le añade una diéresis (*Umlaut*) en alemán (a se convierte en ä, o en ö o u en ü, por ejemplo en el caso de *das Haus* (singular) y *die Häuser* (plural)).

Si bien los ejemplos anteriores no generan problemas de traducción, hay algunos sustantivos que no sufren ningún cambio en plural por ejemplo *der Rechner /die Rechner, der Schüler/die Schüler*. En estos casos el traductor solo puede detectar con ayuda del verbo si se está haciendo referencia a la forma singular o plural del sustantivo y esto puede generar problemas y por consiguiente errores en la traducción.

En el plano morfológico también son interesantes en alemán las nominalizaciones de verbos a través del sufijo *-ung*. Este tipo de nominalizaciones pueden describir un evento, un estado o un objeto. En los casos en que la nominalización va precedida por una preposición o un adverbio como *nach*; las nominalizaciones pueden interpretarse de diferentes maneras y solo el contexto puede eliminar esta ambigüedad (Eberle et al. 2008, 38). Eberle trabaja con un corpus alemán-inglés, por eso incluimos aquí las traducciones al castellano de algunos de los ejemplos que utiliza el autor en el artículo citado. Las traducciones al castellano las hemos realizado especialmente para estos ejemplos. A continuación damos dos ejemplos:

	Funciona como...	Ejemplo y traducción*
<i>nach</i>	un adverbio temporal	<i>Die 1. Wiederholungsprüfung muss nach Mitteilung des Prüfungsergebnisses zum nächsten regulären Prüfungstermin abgelegt werden.</i>

1. Traducción, traducción técnica, problemas de la traducción

	Funciona como...	Ejemplo y traducción*
		La primera repetición del examen debe realizarse en la siguiente fecha de examinación posible, después de que se hayan publicado los resultados del examen.
	una preposición	<i>Nach Pierre's Darstellung des Unfalls, war der Fahrer betrunken.</i>
		Según la descripción de Pierre del accidente, el conductor estaba ebrio.

Tabla 42: Nominalización con *-ung* y preposiciones

1.6 Conclusiones del capítulo

Las diferentes clasificaciones de problemas y errores mencionadas tienen algo en común, para todos los autores son conceptos importantes para la evaluación de la calidad de la traducción. En algunos casos describen el mismo problema o error, solo que con una denominación diferente. La clasificación de Schmitt, quien analiza los problemas de las traducciones técnicas, es importante para nuestra tesis debido a la temática misma, si bien la clasificación de errores que propone tiene en cuenta básicamente el texto original.

La conclusión de este capítulo, siguiendo la orientación de Hurtado (2011, 305) es que todas las clasificaciones o tipologías de errores nos proporcionan un metalenguaje que nos permite el poder hablar de ellos, pero no dan cuenta por si mismas de la gravedad o repercusión del error, ni pueden explicar porque se produce éste mismo. Habiendo listado algunos de los tipos de problemas que se presentan en la combinación lingüística alemán-castellano, podemos afirmar que los problemas más frecuentes que encontramos y pueden ocasionar errores son los que surgen por la ambigüedad sintáctica o semántica. Pero también vimos como existen diferencias lingüísticas de orden morfo-sintáctico en la combinación de idiomas de nuestra tesis que pueden ocasionar problemas de traducción.

Con una la clasificación general de los problemas y viendo qué problemas encontramos en la combinación lingüística de nuestra tesis que nos interesa nos resta analizar qué tipo de problemas surgen cuando traducimos utilizando motores de traducción automática. Antes de hablar de los problemas; pero en primer lugar introduciremos el tema de la traducción automática en general y de los diferentes métodos y sistemas que se emplean en este tipo de traducción.

2. Tecnologías de la traducción, traducción automática y sus problemas

El ejercicio de la traducción ha cambiado en las últimas décadas y la utilización de ordenadores no es algo ajeno a la tarea traductora. En este capítulo nos referiremos particularmente a la traducción automática, como uno de los últimos desarrollos de las tecnologías de la traducción.

Existen diferentes clasificaciones de las tecnologías de la traducción, una de las primeras y más conocida es la que proponen Hutchins y Somers (1992, 148). Según estos autores —dependiendo de la intervención de la tecnología en el proceso de traducción y de la participación del traductor humano— se pueden distinguir diferentes niveles de automatización que van desde la traducción completamente automática de alta calidad (FAHQT o *Fully Automatic High Quality Translation*), hasta la traducción tradicional (*Traditional Human Translation*). Dependiendo del papel que la tecnología tenga en la traducción, también se puede distinguir entre *Human Aided Machine Translation* (HAMT) y *Machine Aided Human Translation* (MAHT). A diferencia de la FAHQT o la *Fully Automated Machine Translation* (FAMT) en los procesos de HAMT se requerirá, en caso de ambigüedad de la intervención humana. También podemos hablar del concepto de *Machine Aided Human Translation* (MAHT) como el tipo de traducción en la que el ser humano realiza la mayor parte del trabajo, pero utiliza algunos sistemas informáticos —por ejemplo diccionarios o correctores ortográficos— como ayuda para realizar la traducción (Maegaard 2001).

De acuerdo con la definición de Hutchins y Somers la traducción asistida por ordenador o TAO se ajusta a las categorías de HAMT o MAHT pero compartimos con Martín-Mor la idea de que esta distinción es más un asunto teórico que práctico, ya que la línea de separación entre estos tipos es cada vez más difusa (2011, 90).

2. Tecnologías de la traducción, traducción automática y sus problemas

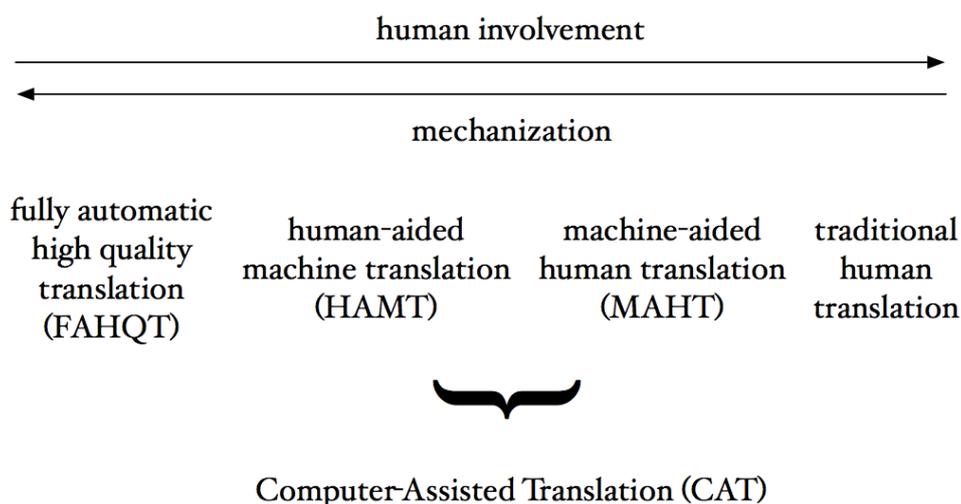


Ilustración 1: Grados de automatización del proceso de traducción según Hutchins y Somers(1992, 148)

Con los nuevos avances tecnológicos como la virtualización de procesos en la nube y la utilización creciente de la traducción automática nos encontramos con nuevas tareas (como la edición previa, la gestión de procesos de traducción automática y la posesición) y por ello mismo son posibles otro tipo de clasificaciones como la de Alcina (2008) que identifican cuatro criterios adicionales para la clasificación de la automatización en la traducción. Según esta autora las tecnologías de la traducción se pueden clasificar dependiendo de la fase del proceso en que se emplean; según el nivel de conocimientos tecnológicos necesarios; según su relación con la traducción y por último según su dimensión (didáctica, profesional o de investigación).

Las clasificaciones de las tecnologías de la traducción se refieren, aunque sea de manera implícita, también al proceso de traducción. Investigadores del grupo Tradumàtica (Martín-Mor, Pique-Huerta, & Sánchez-Gijón, 2015) propone una división del proceso de traducción en fases y tareas que se llevan a cabo con herramientas. Los autores no tratan de describir un proceso único de traducción, sino de identificar algunas de las tareas que se dan en los diversos procesos digitalizados de la traducción y agruparlas por fases. Dicho modelo es aplicable a la traducción técnica y enumera tareas que no necesariamente se dan en cada encargo de traducción.

2. Tecnologías de la traducción, traducción automática y sus problemas

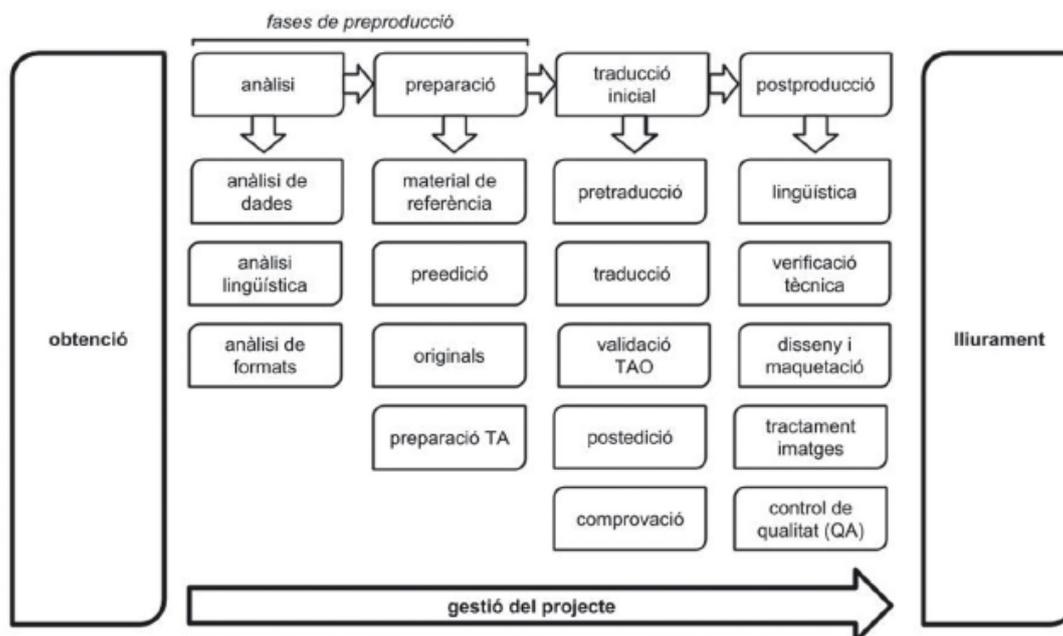


Ilustración 2: Fases y tareas del proceso de digitalización de la traducción (Adrià Martin-Mor, Pique, y Sánchez-Gijón 2016, 34)

Según este modelo los procesos de la digitalización pueden agruparse en tres fases, cada una con determinadas tareas y durante las cuales se pueden utilizar determinadas herramientas: 1) La fase de preproducción que a su vez abarca la fase de análisis y preparación del proceso de traducción, en la que se incluyen las que tienen que ver con la toma de decisiones necesarias para realizar el encargo de traducción (gestión de formatos, elaboración de presupuesto, planificación), es decir abarca las tareas que permiten identificar qué recursos técnicos y humanos se requieren para realizar el encargo. En esta fase nos encontramos también con las tareas que tienen que ver con la preparación del proyecto, por ejemplo, selección del material de referencia, preparación de originales (preedición), organización del trabajo, selección de memorias de traducción, definición de procesos previos. 2) En la fase de traducción se agrupan todas las tareas necesarias para la obtención de una primera versión de la traducción como la pretraducción, la traducción misma y la validación de las propuestas provenientes de una memoria de traducción. También en el caso que se utilice traducción automática se incluirá aquí la posesión de las propuestas. 3) La fase de postproducción es la etapa final de elaboración del producto traducido optimización que incluye las tareas como la revisión lingüística y técnica de la primera versión traducida, esta fase también incluye la aplicación de sis-

2. Tecnologías de la traducción, traducción automática y sus problemas

temas de control de calidad. Una vez concluida esta fase se puede realizar la entrega del texto final del proyecto de traducción.

Interesante en este modelo es la inclusión de la traducción automática y la posesición en el proceso de traducción que permite ver como hay una hibridación en dicho proceso: no solo nos encontramos el empleo de la traducción automática en la fase de producción, sino que también podemos tener el caso de que el texto origen no sea producido por humanos, sino sea parcial o totalmente una traducción automática. Pero antes de entrar al tema de la traducción automática y la posesición nos referiremos primero al concepto de traducción asistida por ordenador (TAO).

2.1 Traducción asistida por ordenador

A mediados de la década de los 80 se empiezan a desarrollar las llamadas herramientas TAO (Traducción Asistida por Ordenador) o CAT (*Computer-Aided Translation* o *Computer-Assisted Translation tools*). Bajo este nombre se agrupa una serie de recursos informáticos que hoy en día se utilizan en muchos procesos de traducción. Estos recursos o herramientas incluyen obras de referencia en línea (diccionarios monolingües, bilingües), correctores ortográficos y gramaticales, procesadores de textos, etc.

La característica principal de una herramienta TAO es que no realiza traducciones, sino las reutiliza. Es decir, las herramientas TAO son sistemas de recuperación de información (Reinke 2003). La traducción la realiza un traductor y la guardada en el sistema (original y su traducción) para usarla como propuesta de traducción en el futuro. Como herramientas TAO se consideran las memorias de traducción, los gestores terminológicos, las herramientas de alienación de texto, las herramientas de localización de software o de páginas web. Algunos autores incluyen aquí también las herramientas de subtítulos y de ayuda lexicográfica o de explotación de corpus que permite buscar concordancias, colocaciones o agrupaciones de palabras (Bermúdez 2014, 131)

El software, para ser exactos, el paquete de software que se utiliza para traducción asistida por ordenador se denomina de diferentes maneras, debido en parte a que se toman los nombres en inglés. Es por esto que hablamos de herramientas TAO, también denominadas Sistemas de Gestión y de Edición de Traducciones (Adrià Martin-Mor, Pique, y Sánchez-

2. Tecnologías de la traducción, traducción automática y sus problemas

Gijón 2016, 14) y de *Translator's Workstation* o estaciones de trabajo del traductor. Estas estaciones de trabajo están integradas por varios programas, a menudo independientes, que los traductores instalan y configuran a su gusto en el propio ordenador. Cada vez más, sin embargo, se tiende hacia entornos de trabajo compuestos por varias herramientas que permiten al profesional de la traducción tener acceso a los recursos que necesita sin tener que cambiar de entorno. Un ejemplo de esta tendencia lo encontramos en el *Google Translator Toolkit* que incorpora un editor de texto, una memoria de traducción, un módulo de traducción automática, un corrector ortográfico y otras funciones adicionales como la posibilidad de chatear con otros traductores que estén en línea (Adrià Martin-Mor 2011, 68).

Como mencionamos anteriormente los sistemas de gestión de memorias de traducción tienen diversos componentes, que dependen del producto respectivo. Sin embargo existen ciertas partes esenciales que se encuentran en la mayoría de los sistemas: la memoria de traducción o *translation memory*, el editor, el filtro de formatos, un componente para alineación de traducciones, una base de terminología y un módulo para el manejo y administración de proyectos. En la actualidad la mayoría de los sistemas han integrado un componente para la traducción automática.

Un término con el que nos encontramos más frecuentemente cuando nos referimos a herramientas TAO es la memoria de traducción (*TM: translation memory*). Las memorias de traducción son unos de los sistemas más utilizados actualmente por los traductores profesionales. Pero en realidad con el concepto de memoria de traducción se está tomando solo un componente para la identificación de todo un paquete de software, si bien las memorias de traducción son el componente principal de los sistemas de gestión de memorias de traducción (SGET).

Los sistemas de gestión de memorias de traducción permiten el almacenamiento de fragmentos de textos originales asociados a sus respectivas traducciones con el objetivo de utilizarlos de nuevo en el futuro. Un concepto clave de las memorias de traducción es el de la segmentación, es decir, la división del texto a traducir en fragmentos de longitud limitada, por ejemplo gracias a marcas tipográficas como saltos de párrafos, puntos y signos de interrogación o exclamación. El que los segmentos sean limitados aumenta la posibilidad de su utiliza-

2. Tecnologías de la traducción, traducción automática y sus problemas

ción en el futuro. Adicionalmente pueden guardarse con cada segmento meta datos como autor, fecha de producción o modificación, número de repeticiones, etc. (Adrià Martin-Mor 2011, 71).

Al traducir con un SGET nos encontramos con segmentos que no tienen traducción en la memoria o con coincidencias que pueden ser exactas —es decir que el segmento del texto original coincide 100% con uno ya traducido y guardado en la memoria— o parciales, cuando solo algunas partes del segmento coinciden o el formato y/o la posición de las palabras es diferente. Muchos sistemas de gestión de memorias de traducción permiten al usuario elegir el umbral de coincidencia por debajo del cual no se deben mostrar coincidencias parciales, de esta manera se evita que se visualicen propuestas poco relevantes.

Hay tres maneras de obtener una memoria de traducción (Somers, 2003a, p. 33): en primer lugar creando una memoria misma, es decir, aumentando el contenido de manera interactiva con la ayuda de un SGET. En este caso, el programa segmenta automáticamente el texto original y a medida que el traductor introduce las traducciones almacena los segmentos. Este proceso se denomina traducción interactiva (Bowker 2002, 109). La segunda manera es importar una memoria ya existente. Esta manera es empleada por muchas empresas de traducción y ha sido facilitada por la aparición del estándares de intercambio de memorias como TMX (TMX Translation Memory eXchange), permiten importar una memoria de traducción en un sistema diferente al en que haya sido creada. Los sistemas de gestión de memorias de traducción también permiten exportar los contenidos de la base de datos en un formato de intercambio, así como también permiten exportar proyectos en el formato XLIFF (XML Localization Interchange File Format). Finalmente se puede obtener una memoria de traducción mediante la alineación de textos originales con sus respectivas traducciones. Este proceso permite alimentar una memoria de traducción con segmentos de traducciones hechas en el pasado sin utilización de SGET. Los sistemas más difundidos de gestión de memorias incorporan un módulo para la alineación de textos, que permiten al usuario supervisar la alineación automática realizada por el programa y modificarla de manera interactiva en caso que sea necesario.

2. Tecnologías de la traducción, traducción automática y sus problemas

Los SGET también ponen a disposición de los usuarios editores, en los que se realiza la traducción, y donde el traductor puede utilizar la memoria, aceptando una propuesta grabada en ella o adaptado un *fuzzy match* (coincidencia parcial). Los cambios y las nuevas traducciones realizadas en el editor se guardan en la memoria para futuras traducciones. A estos componentes elementales se han añadido otros como aquellos que permiten utilizar y trabajar con una base de datos de terminología, administrar diversas tareas o proyectos de traducción y tareas administrativas como alinear textos ya traducidos e importar las traducciones u otras memorias y como enunciamos anteriormente algunas herramientas TAO ofrecen también la posibilidad de ver propuestas de traducción automática.

Una ventaja adicional para los traductores es que los editores de estos sistemas están en capacidad de leer múltiples formatos de texto, de tal manera que el traductor no tiene que tener los programas en que se han editado los originales, sino solo leerlos en la aplicación TAO. La mayoría de los editores funcionan hoy en día como un procesador de texto, es decir, puede hacer correcciones ortotipográficas, permiten búsqueda de textos y ajustes de formatos.

Los sistemas comerciales más conocidos son SDL Trados, Déjà Vu, MemoQ, MetaTaxis, STAR Transit, Wordfast y Across; pero también existen aplicaciones de código abierto como Anaphraseus, OmegaT, Omegat+, Aushumatoite, OPEN TM2, Olifant, Virtaal, Poedit y Tikal (Ortiz Jimenez 2014).

Aunque las herramientas TAO modernas son capaces de hacer algunos ajustes de texto, como por ejemplo adaptar los números, ciertas etiquetas o la puntuación, estas sustituciones no pueden definirse como nuevas traducciones, lo que si sucede —como veremos— cuando se emplea un sistema de traducción automática. Otro punto de diferencia entre la traducción automática y las herramientas TAO es que en éstas últimas es el traductor quien decide si una propuesta de traducción puede ser aceptada o no y quien hace, en caso necesario, la nueva traducción (Harold Somers 2003b).

2.2 *La traducción automática*

La investigación en el área de la traducción automática se inició hace más de 60 años impulsada por la necesidad estratégica durante la guerra fría de conocer lo que se escribía en otros idiomas, particularmente en ruso. Debido a la falta de resultados satisfactorios se redujo

2. Tecnologías de la traducción, traducción automática y sus problemas

la financiación de la investigación en esta área y solo a partir de los 80s —a partir del aumento del volumen de datos que deben ser traducidos— ha vuelto a renacer el interés, en primer lugar por parte de la industria, por apoyar la investigación en esta área.

En este capítulo no haremos un resumen diacrónico de la evolución de la traducción automática sino que nos referiremos a ciertos términos que utilizaremos durante toda la tesis y cuya comprensión es básica para el desarrollo de los capítulos siguientes.

El auge de la utilización de los sistemas de traducción automática en los últimos tiempos se debe particularmente a que permiten traducir más palabras en menos tiempo y, lo que es más importante para el mercado, utilizando menos recursos humanos. Existen tres sistemas principales de traducción automática: aquellos basados en reglas RBMT (*Rule-based machine translation*); los sistemas basados en datos o corpus que pueden ser estadísticos SMT (*Statistical Machine Translation*) o basados en ejemplos EBMT (*Example Based Machine Translation*). Y finalmente los sistemas de traducción automática basados en la generación de sistemas neuronales NMT (*Neural Machine Translation*).

Antes de referirnos a los problemas de la traducción automática describiremos brevemente como funcionan dichos modelos y motores de traducción automática, cómo se valora y estima la calidad de las traducciones realizadas de esta manera. También describiremos algunos de las herramientas que se han desarrollado para la evaluación de la calidad de la traducción automática, así como los métodos de medición de errores.

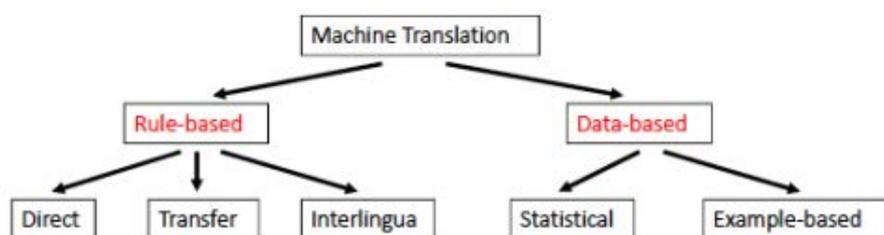


Ilustración 3: Modelos de traducción automática (Lavanchy 2010, 2)

Existe una interpretación de la traducción automática como una tarea de traducir las oraciones de un texto de forma independiente. Traducir frases u oraciones aisladas implica,

2. Tecnologías de la traducción, traducción automática y sus problemas

por un lado, la construcción de una interpretación del significado de la oración en el lenguaje origen o base (a traducir) y, en segundo lugar, la construcción de una oración en el lenguaje meta según la interpretación que se hizo. Para construir la interpretación de la oración se comienza por determinar o definir las palabras como bloques de construcción, que se pueden agrupar posteriormente en grupos más complejos. Para este proceso no es necesaria la participación de traductores profesionales que interpreten o entiendan plenamente las frases del texto, siempre y cuando se puedan transformar las estructuras o modelos sintácticos de las frases y traducir las palabras que las forman (prestando especial atención a la terminología). Esta afirmación se sustenta en un principio de la semántica composicional según el cual la interpretación de una frase está completamente determinada por la interpretación de las palabras que la componen y por la estructura que las agrupa. No es lo mismo “Ellos escriben cartas” que “Ellos escribieron artículos” o “Israel amenaza a Palestina” que “Palestina amenaza a Israel”.

Otra premisa en la que se basa la traducción automática es la idea de que no es necesario construir una interpretación completa de una frase para traducirla, basta con transformar la estructura de la sentencia en el lenguaje base en una estructura válida en el lenguaje meta y sustituir las palabras de la oración por los equivalentes adecuados en el lenguaje meta. Es precisamente aquí donde los diferentes sistemas de traducción utilizan diferentes tipos de interpretaciones: los sistemas basados en reglas y los sistemas basados en corpus y estadísticas. Los sistemas basados en reglas se apoyan principalmente en la transferencia sintáctica y aplican reglas obtenidas gracias a un análisis previo a la representación del texto original. Así se obtiene una representación intermedia del texto en la lengua meta a partir de la que se genera la traducción. Para ejecutar dicha tarea los sistemas emplean secuencias de formas léxicas, árboles de análisis sintáctico, árboles de dependencias e interlingua.(Sánchez-Martínez 2012, 150). A continuación entraremos más en detalle en la descripción de estos sistemas.

2.2.1 *Sistemas de traducción automática basados en reglas*

La investigación en el área de la traducción automática se concentró al comienzo en sistemas basados en reglas que emplearon técnicas tales como la traducción directa, la transferencia y la traducción basada en una lengua interlingua (Koehn 2010, 15-16).

2. Tecnologías de la traducción, traducción automática y sus problemas

El acercamiento más tradicional a la traducción automática es de carácter deductivo; es decir, se trata de transferir al sistema o motor de traducción el conocimiento lingüístico por medio de reglas. Es por ello que los modelos desarrollados a partir de este acercamiento se denominan “Sistemas de traducción automática basados en reglas”. Estos sistemas se apoyan en la idea de que la tarea fundamental de la traducción es encontrar la equivalencia semántica. Tal equivalencia se basa en la idea de que todos los lenguajes del mundo comparten una misma estructura lógica (Arnold et al. 1994, 30ss). Por lo tanto, si se conoce y domina dicha estructura, se podrá traducir cualquier idioma a otro.

Dentro de los sistemas de traducción automática basados en reglas encontramos diferentes métodos de resolver el problema de la equivalencia conceptual. Se habla de un método directo si se analizan los dos idiomas implicados directamente y se habla de métodos indirectos si se trata de resolver el problema de la equivalencia desarrollando un tipo de representación o representaciones intermedias. Para estos métodos indirectos es imprescindible tener información tanto de la lengua de origen como de la lengua de destino o meta y también de una representación intermedia que capture el significado de la oración base para poder generar la oración meta (A. Martin-Mor 2017).

Traducción directa

Es el sistema más simple es traducir palabra-por-palabra y luego a través de un componente sintáctico ajustar o reordenar las palabras según la estructura del idioma meta. La traducción se realiza en un solo paso, se consulta el significado de las palabras de origen en un diccionario y directamente se generan sus equivalentes en la lengua meta. Posteriormente se aplican las reglas de flexión, concordancia y reordenamiento a través de las cuales se pretende solucionar los problemas principales de la traducción.

En este método no se tiene en cuenta la estructura semántica de la frase, ni las relaciones semánticas existentes entre las palabras que la comprenden. La mayor parte de los resultados son solo parcialmente utilizables en escenarios especiales, ya que con este sistema no se puede garantizar la traducción correcta, por ejemplo, de oraciones ambiguas o idiomáticas (Stein 2009, 8). Un ejemplo de este tipo de aproximaciones lo encontramos en el proyecto *Georgetown Automatic Translation-Project GAT*, en el que se basa el sistema comercial SYSTRAN (Hutchins 2001, 434).

2. Tecnologías de la traducción, traducción automática y sus problemas

El modelo de transferencia

El aspecto característico del método de transferencia es que propone dos representaciones intermedias, una por cada idioma de la combinación a traducir. La traducción se realiza en tres fases: primero el análisis, luego la transferencia y por último la generación. La transferencia se puede realizar tanto a nivel léxico, como en el sintáctico o el semántico.

Como primera medida se realiza un análisis morfológico del texto original para identificar sus componentes y clasificarlos según su función; luego se realiza categorización léxica en la que se analiza el contenido y se resuelven las ambigüedades. Después de esto se lleva a cabo la transferencia léxica en la que a cada palabra del texto original se le asigna un equivalente en el idioma meta. Después de que se ejecuta esta fase se realiza la transferencia estructural, es decir, se analiza el texto no a nivel de las palabras, sino teniendo en cuenta segmentos mayores, con el fin de encontrar expresiones que requieran un tratamiento especial, por ejemplo, refranes, metáfora, etc. Cuando se culmina este análisis el generador morfológico se encarga de asociar cada elemento identificado en las etapas anteriores con un equivalente en la lengua meta (Reimann 1991, 64).

Los sistemas de traducción automática que se enfocan en la transferencia son Apertium (Forcada, Ginestí-Rosell, y Nordfalk 2011) y Lucy Translation.

Con este acercamiento es posible obtener traducciones de mejor calidad y es por esto que los sistemas de transferencia han tenido auge durante las últimas décadas. Sin embargo el problema principal de este sistema es que se pueden pasar por alto diferencias en el orden gramatical de las oraciones que pueden tener repercusiones sintácticas y por ello funciona mejor en lenguas estrechamente relacionadas y con estructuras similares (Hutchins y Somers 1992, 128).

El modelo de interlingua

Los sistemas de traducción basados en este modelo realizan su función en dos pasos, en primer lugar se le asigna una estructura al texto origen (una oración en el lenguaje 'intermedio' que contiene el significado del texto a traducir), y posteriormente partiendo de esta estructura se genera el texto meta sin tener en cuenta cual es el lenguaje de origen. La información se-

2. Tecnologías de la traducción, traducción automática y sus problemas

mántica suele estar recogida en una base de conocimientos a la que se puede acceder durante el proceso de traducción.

La traducción en una o de una interlingua solo se tiene que realizar una vez para cada idioma nuevo. Según este modelo si a x-idiomas se debe añadir uno nuevo, tendrán que hacerse 2x direcciones de traducción. Por ejemplo, cuando aumentaron de 11 a 20 los idiomas de la UE, se necesitaron en vez de 110, 380 pares de traducción.

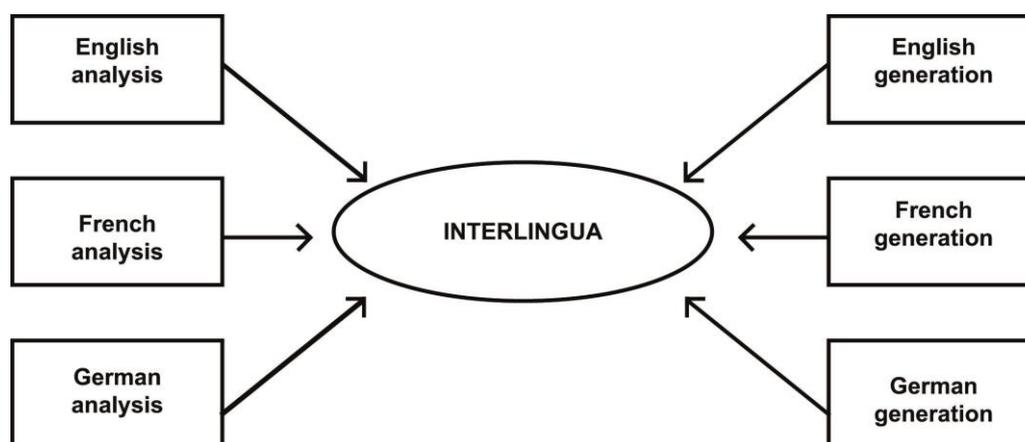


Ilustración 4: Modelo basado en la interlingua con seis pares de lenguas (Hutchins y Somers 1992, 74)

El idioma intermedio tiene que estar granulado finamente, para que se puedan contemplar todas las diferencias en todos los idiomas. Esto implica en el proceso de traducción un volumen de traducción cada vez mayor, en especial de los idiomas que son familiares.

La transferencia sintáctica es muy compleja debido a las diferencias en las construcciones sintácticas de los idiomas. Entre los proyectos que han utilizado este método se encuentran ATLAS (Fujitsu) y PIVOT (NEC). Un proyecto japonés que continúa en la actualidad es UNL (Universal Networking Language, de la United Nations University). En EEUU se han realizado varios proyectos basados en interlinguas: KBMT (Carnegie Mellon University), ULTRA (New Mexico State University), PANGLOSS (CMU, NMSU y University of Southern California), UNITRAN (University of Maryland), todos estos son proyectos de investigación académicos y ninguno de ellos se ha comercializado. En Europa se pueden reseñar dos proyectos que se inspiran en el modelo interlingua: DLT (de la empresa holandesa BSO) que utiliza el esperanto como interlingua y ROSETTA (PHILIPS) basada en la gramática de Montague.

2. Tecnologías de la traducción, traducción automática y sus problemas

Una forma común de visualizar los tres enfoques de las arquitecturas de sistemas pre-estadísticos es el triángulo de Vauquois. Bernard Vauquois (1975) fue un líder influyente en la investigación en traducción automática en la década de 1970 en la Universidad de Grenoble, uno de los centros de investigación más importantes de traducción automática basada en reglas. En tanto se asciende en el triángulo de Vauquois se mantiene los lemas de las palabras con contenido léxico, es decir, nombres, verbos, adjetivos y adverbios; la información morfológica de las palabras se convierte en rasgos: género, número, tiempo, aspecto, función, etc.; se eliminan las palabras funcionales o gramaticales como artículo, conjunciones y su información también se convierte en rangos: +definido / -definido, o función: sujeto, objeto.

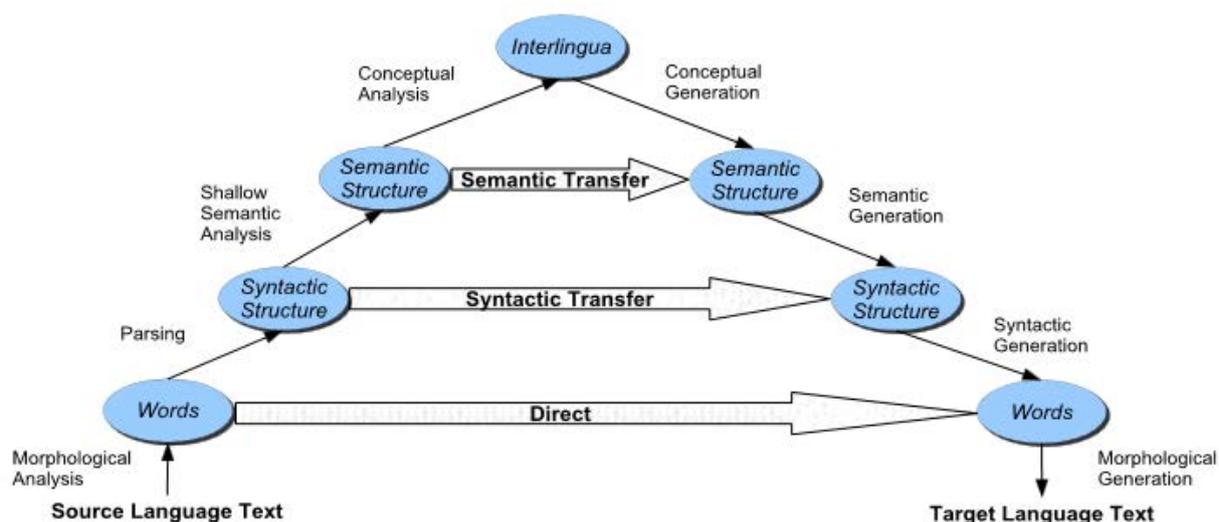


Ilustración 5: Triángulo de Vauquois (Van Gompel 2009, 16)

2.2.2 Sistemas basados en datos o corpus

Los sistemas que se sustentan en datos o corpora se diferencian en aquellos que se basan en ejemplos y de los que se basan en estadísticas. La traducción automática basada en estadísticas ha sido más desarrollada y ha experimentado muchos cambios desde finales de los 80 y comienzos de los 90. A continuación describiremos los rasgos principales de estos sistemas.

2.2.2.1 Sistemas de traducción automática estadística

Los sistemas basados en estadísticas parten de la premisa de que al traducir una oración de un idioma origen ' e ' a un idioma meta ' f ', todas las oraciones del idioma meta ' f ' son potencialmente la traducción correcta de la oración en cuestión ' e ' en el idioma origen. Sin

2. Tecnologías de la traducción, traducción automática y sus problemas

embargo algunas oraciones tienen una probabilidad más alta de ser la traducción correcta. Para poder encontrar esta probabilidad el sistema utiliza textos bilingües o corpora alineados y mediciones de probabilidad.

Para poder realizar el cálculo probabilístico los sistemas tienen que ser entrenados y es para ello que se utilizan traducciones existentes (textos bilingües, también conocidos como corpus paralelos). Dichos textos deben ser en primer lugar segmentados en frases con ayuda, por ejemplo, de los signos de puntuación y de la información de formato. En segundo lugar las oraciones o segmentos encontrados deben ser alineados, es decir, se deben identificar las traducciones y sus correspondientes oraciones en el texto original. La alineación puede ser 1-1 o 1-N o N-1 (Koehn 2010: 56).

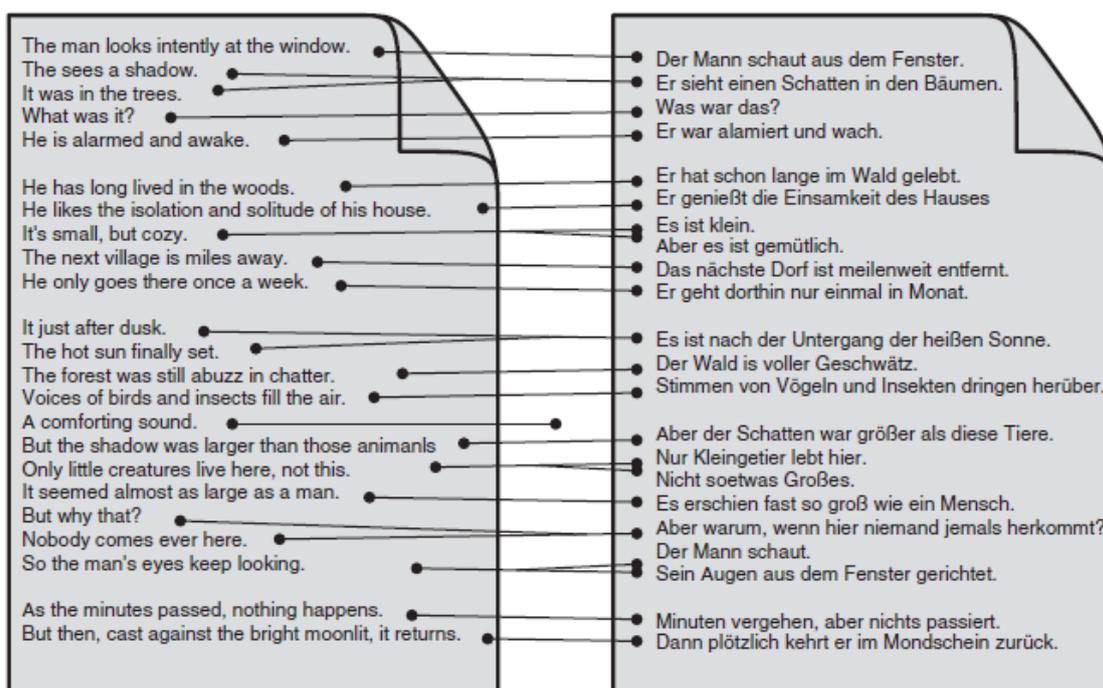


Ilustración 6: Alienación según Koehn (2010, 56)

En el proceso de alineación se parte de que todas las combinaciones o alineaciones de palabras son igualmente probables. Al comparar con el corpus se puede deducir cuáles son las alineaciones más probables y así revelar la estructura “oculta” de alineación. Con el alineamiento de palabras se puede obtener fácilmente un diccionario bilingüe de probabilidades. Este procedimiento basado en el cálculo de probabilidades es conocido como **modelo de tra-**

2. Tecnologías de la traducción, traducción automática y sus problemas

ducción palabra por palabra. En el modelo de traducción una oración en el lenguaje meta T es tanto más fiel al original S , entre más alta sea la probabilidad de que T sea la traducción correcta de la oración S . La probabilidad $p(s/t)$ se calcula con base en la frecuencia en que se encuentra en un corpus bilingüe alineado.

Como $p(e)$ se formula la probabilidad apriori de que 'e' suceda. Por ejemplo, si 'e' es la cadena en inglés *I like snakes*, entonces $p(e)$ es la probabilidad de que una determinada persona en un momento determinado va a decir *I like snakes* en lugar de decir otra cosa.

Con $p(f|e)$ se describe una probabilidad condicional. La posibilidad de que a la cadena 'f' le corresponda (e). Si en nuestro ejemplo (e) es *I like snakes* y (f) en francés es *maison bleue* la posibilidad de que al ver (e), un traductor escriba (f) sería = 0. Con $p(e,f)$ se formula una probabilidad conjunta, es decir la posibilidad de que (e) y (f) se den al mismo tiempo. Si (e) y (f) no tienen ninguna relación la fórmula se escribe de la siguiente manera: $p(e,f) = p(e) * P(f)$.

Si tenemos una oración en francés (f) y queremos buscar la traducción más probable en inglés (e), es decir, $p(f|e)$, entonces se escribe la siguiente formula:

$$\underset{e}{\text{argmax}} = p(f|e)$$

Aquí se resuelve *argmax* de la siguiente manera: la sentencia (e) en inglés, que entre todas las frases, obtiene el valor más alto de probabilidad o sea $P(e|f)$.

Para calcular la probabilidad se utilizan algoritmos o reglas como la de Bayes. El teorema de Bayes es un caso especial de probabilidad condicional que se aplica cuando se desea calcular la probabilidad condicional de un evento que ocurrió primero, dependiendo de lo que ocurrió después:

$$p(x|y) = \frac{p(y|x) p(x)}{p(y)}$$

Ilustración 7: Formula de Bayes (Koehn 2010, 69)

2. Tecnologías de la traducción, traducción automática y sus problemas

En esta fórmula “*x*” sería la traducción e “*y*” el texto o palabra a traducir. Si se utiliza siempre el equivalente más probable proporcionado por un diccionario lingüístico de probabilidades se traduce palabra por palabra y no se tiene en cuenta el contexto. Ello puede tener como resultado que las traducciones no fluyan y no tenga sentido. Para alcanzar inteligibilidad y fluidez se requiere del **modelo del lenguaje meta**. En el modelo del lenguaje una traducción *T* es mejor en cuanto mayor sea la probabilidad de que se utilice realmente en el lenguaje meta. Mientras que en el modelo de traducción se utilizaba un corpus bilingüe para hacer el cálculo de la probabilidad en este modelo se calcula la probabilidad de que aparezca la oración en un corpus del lenguaje meta, es decir monolingüe. Este modelo permite medir la probabilidad de que la traducción automática de una oración o segmento sea una oración correcta y fluida en el lenguaje meta, en otras palabras, la probabilidad estadística nos informa aquí que tan aceptable o usual es una oración determinada en el lenguaje meta. Ya que hacer una búsqueda de un texto complejo puede ser muy extensa y no dar los resultados esperados, se rompe el cálculo de la probabilidad de una larga cadena en partes más pequeñas, para las cuales se pueden recopilar suficientes estadísticas. El método más común en el modelo del lenguaje es el uso de **n-gramas**. Un *n-grama* es una subsecuencia de ‘*n*’ elementos en una secuencia dada, es decir ‘*n*’ es el número de palabras, por ejemplo 3 —trigramas— (Koehn 2010, 95). Con este tipo de análisis se pretende calcular la probabilidad de que a una secuencia de letras o palabras le siga otra letra o palabra. Este tipo de cálculo estadístico es importante porque funciona en cada idioma o alfabeto y además es de gran utilidad si se tiene a disposición una cantidad de datos considerable, que de otra manera sería imposible de analizar. Durante el análisis se puede observar que es más probable que a una combinación como *the red* le siga la palabra *house* que a la opción inversa. Es decir, es más probable encontrar la combinación *the red house* que *house the red* (Koehn 2010, 182).

$$p(\text{house}|\text{the red}) = \frac{\text{count}(\text{the red house})}{\text{count}(\text{the red})}$$

La fórmula se puede abstraer de la siguiente manera (Koehn 2010, 183)

$$p(w_3|w_1, w_2) = \frac{\text{count}(w_1, w_2, w_3)}{\sum_w \text{count}(w_1, w_2, w_3)}$$

2. Tecnologías de la traducción, traducción automática y sus problemas

Koehn también nos da un ejemplo con el que se puede entender más claramente cómo funciona este tipo de cálculo. El autor busca en el corpus paralelo de la Unión Europea las palabras que tienen más probabilidad de aparecer después de las secuencias *the green*, *the red* y *the blue*. La probabilidad de que a la secuencia *the green* siga la palabra *paper* es más grande que la probabilidad de que sea la palabra *ecu*; lo mismo que es más probable encontrar la combinación *the red cross*, que solo *the red card* o *the blue box* que *the blue angel*. Por otro lado se puede deducir que si bien la combinación *the red army* no es falsa, no aparece con frecuencia en el corpus seleccionado. Esto implica que aunque este modelo nos da un mejor resultado que los dos modelos explicados anteriormente, no es suficiente aun para generar una traducción correcta.

<i>the green</i> (total: 1748)			<i>the red</i> (total: 225)			<i>the blue</i> (total: 54)		
Word	Count	Prob.	Word	Count	Prob.	Word	Count	Prob.
<i>paper</i>	801	0.458	<i>cross</i>	123	0.547	<i>box</i>	16	0.296
<i>group</i>	640	0.367	<i>tape</i>	31	0.138	.	6	0.111
<i>light</i>	110	0.063	<i>army</i>	9	0.040	<i>flag</i>	6	0.111
<i>party</i>	27	0.015	<i>card</i>	7	0.031	,	3	0.056
<i>ecu</i>	21	0.012	,	5	0.022	<i>angel</i>	3	0.056

Ilustración 8: Comparación de un corpus paralelo (Koehn 2010, 183)

a) Los modelos de IBM

En los *modelos de IBM* (1-5) se combinan el modelo de traducción y el modelo de lenguaje con el fin de mejorar la calidad de la traducción (Koehn 2010, 7–8). El proceso de alienación de palabras en un par de oraciones es probablemente el primer paso que utilizan varios de los motores de SMT. Sin embargo muchas palabras pueden tener varias posibilidades de traducción en otra lengua. Por ejemplo *Haus* en alemán puede ser traducido al castellano como *casa*, cuando se está designando un tipo de edificación o puede tener el sentido abstracto de hogar. El equivalente más probable es que se traduzca *Haus* como ‘*casa*’ y no como ‘*hogar*’. Esta probabilidad viene fijada por un análisis estadístico en el que se calcula cuántas veces aparece traducida la palabra *Haus* como ‘*casa*’ y cuántas como ‘*hogar*’. El método con que se calcula esta probabilidad se denomina *lexical translation probability distribution* (Koehn 2010, 82) y se calcula con la fórmula siguiente:

$$p_{\text{translation model}}(\text{source} \mid \text{target}) \times p_{\text{language model}}(\text{target})$$

2. Tecnologías de la traducción, traducción automática y sus problemas

Podemos calcular cual es la probabilidad de que una oración se traduzca al inglés como el equivalente más probable, pero de acuerdo con el modelo de traducción solo se elige cuando la traducción es fluida. Como su nombre indica esta versión de SMT analiza la formación de los datos de prueba a nivel de las palabras. Esto significa que a una palabra en el idioma origen, debe corresponder una palabra en el idioma meta. Pero a veces también ocurre que una palabra en el idioma origen no solo tiene una palabra correspondiente en el idioma meta y su sentido solo puede ser traducido si se emplean varias palabras en el idioma meta. Por ejemplo en inglés *slap* => en castellano *dar una botifada*. El sistema acepta está posibilidad, pero no en el sentido inverso, es decir no se acepta que a la combinación de varias palabras se les asigne una sola palabra en el idioma meta, *dar una botifada* => *slap*. Otro problema relacionado es que las palabras compuestas no pueden ser analizadas. Esto afecta particularmente a aquellos verbos en idiomas como el alemán cuyos prefijos se separan del lexema en la conjugación *Ich reise morgen wieder ab* (Stein 2009, 11–12).

Como la unidad de traducción es la palabra y el modelo de lenguaje solo se tiene en cuenta el contexto de manera indirecta se presentan problemas no solo al traducir verbos (*go out*) o expresiones compuestas (*as a matter of fact*) sino también colocaciones (*make a decision*). Es por ello que los modelos basados en palabras no son los más desarrollados en los últimos tiempos y únicamente se utilizan para calcular las alineaciones.

b) Traducción automática estadística basada en phrasen

Con el fin de resolver los problemas mencionados anteriormente se han desarrollado nuevos enfoques en la traducción basada en estadísticas. Hoy en día, los sistemas trabajan por lo general a nivel de *phrasen* y no de palabras. Como *phrasen* se denominan en lingüística las unidades sintácticas cerradas, que se contraponen a otro tipo de unidades en las que faltan complementos. La organización de las *phrasen* en estos sistemas no se hace según criterios lingüísticos, lo que aquí se denominan *phrasen* son datos de entrenamiento y prueba del sistema que se dividen en grupos con un tamaño y una secuencia determinada. Este sistema abre la posibilidad de traducir ‘frases’ con varias palabras y además de clasificar previamente ciertas construcciones que podrían generar ambigüedades en la traducción. Por ejemplo con este tipo de sistemas se puede hacer una distinción en la traducción de *pretty* si encontramos la palabra en una oración como *pretty much* o en otra oración como *pretty girl* (Stein 2009, 12).

2. Tecnologías de la traducción, traducción automática y sus problemas

También dan la posibilidad de marcar combinaciones o colocaciones como en el caso a continuación *spass am* → *fun with the*:

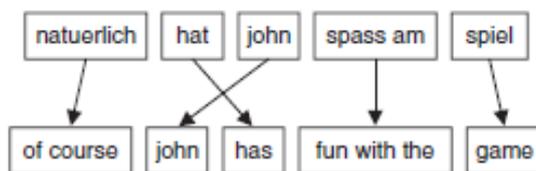


Ilustración 9: Marcaciones de combinaciones en SMT (Koehn 2010, 8)

Este sistema permite no solo que los bloques de construcción sean segmentos de longitud variable, es decir que pueden estar compuestos de una o más palabras. Además permite el tener en cuenta el contexto dentro de la frase y por ello si se tiene como referencia un corpus bilingüe extenso, la traducción será más fiable.

Veamos un poco más en detalle que implican cada uno de los modelos mencionados:

1) *Phrase translation model* (source ↔ target). En este método se utiliza un diccionario probabilístico bilingüe. Funciona igual que un diccionario bilingüe probabilístico basado en palabras pero con segmentos de longitud variable. Para cada par de frases se calculan:

- Los alineamientos posibles de palabras (SMT basado en palabras).
- Se extraen las frases bilingües de los alineamientos de palabras.
- Se calculan las probabilidades según esta fórmula:

$$p(s|t) = \frac{\text{count}(s \leftrightarrow t)}{\text{count}(t)} ; p(t|s) = \frac{\text{count}(s \leftrightarrow t)}{\text{count}(s)}$$

2) Con el *Language model* se hace el cálculo estadístico en el lenguaje meta y se mide qué tan probable es la traducción en dicho idioma.

3) El *Reordering model* se desarrolló para enfrentar el hecho de que en la traducción estadística basada en palabras los modelos de traducción y de lenguaje no generan traducciones fluidas. El problema del reordenamiento sintáctico es uno de los puntos más críticos de la

2. Tecnologías de la traducción, traducción automática y sus problemas

traducción automática ya que existen ciertas combinaciones lingüísticas en las que las estructuras sintácticas son muy diferentes, y aunque solo haya pequeñas diferencias las variaciones pueden ser significativas. Existen diferentes operaciones de reordenamiento léxico posibles:

- Monótona: si no es necesario realizar ninguna reordenación

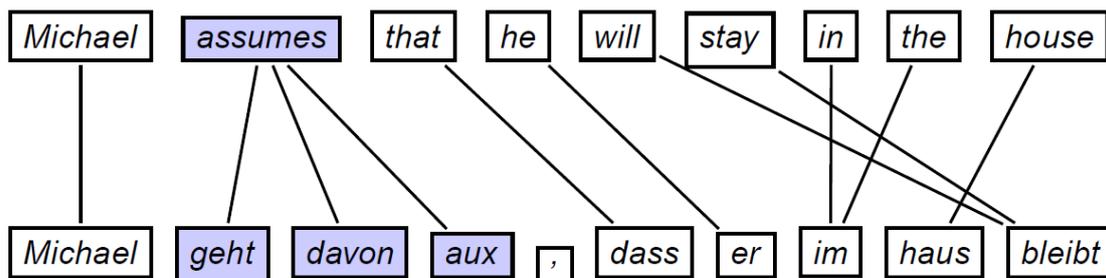


Ilustración 10: Tipo de ordenamiento monótono (Sánchez-Martínez 2014)

- Intercambio: si la 'frase' (como bloque, no como oración) se intercambia con una anterior. Este es el caso que se da cuando en un idioma se cambia la posición por ejemplo de un adjetivo o frase con función de complemento circunstancial o por la posición del verbo.

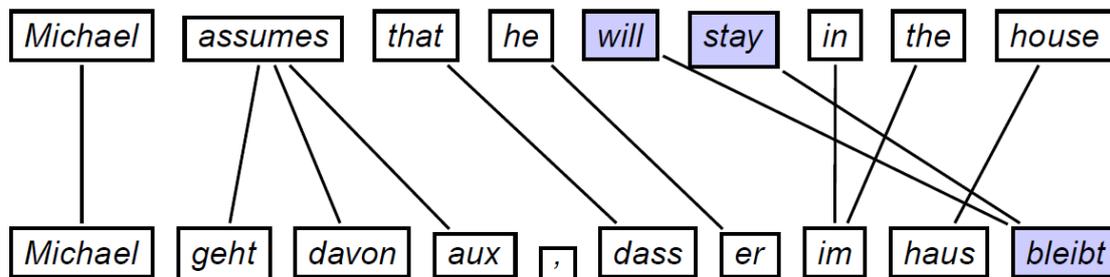


Ilustración 11: Tipo de ordenamiento discontinuo (Sánchez-Martínez 2014)

- Discontinua: si el reordenamiento no es ni monótono, ni de intercambio

Para el cálculo del reordenamiento se utiliza la siguiente formula:

$$p(\text{reordering operation}|s,t) = \frac{\text{count}(\text{reordering operation}, s \leftrightarrow t)}{\text{count}(s \leftrightarrow t)}$$

2. Tecnologías de la traducción, traducción automática y sus problemas

4) Con el *Lexical weighing model* se trata de solucionar el que algunos pares de frases poco frecuentes pueden causar problemas, sobre todo si se recogen a partir de datos que contienen ruido. Con ayuda de un diccionario bilingüe probabilístico basado en palabras se puede medir qué tan confiable es una frase bilingüe. Esto es muy importante porque el modelo de traducción de frases no es fiable para frases poco frecuentes o largas (Koehn 2010, 139).

¿Cómo podemos juzgar si la traducción de una frase poco frecuente es fiable? Si fragmentamos una frase en las traducciones de las palabras que la componen, podemos comprobar si coincide y qué tanto coincide. El modelo de ponderación léxica es básicamente un método de perfeccionamiento. La ponderación léxica se base en las distribuciones de probabilidad y por ello se requiere de suficientes datos estadísticos para poder garantizar estimaciones de probabilidad más fiables.

Figure 5.7 Lexical weight p_w of a phrase pair (\bar{e}, \bar{f}) given an alignment σ and a lexical translation probability distribution w : Each English word has to be explained by foreign words using the distribution w . If aligned to multiple foreign words, the average is taken. If unaligned, $w(e_i|NULL)$ is factored in.

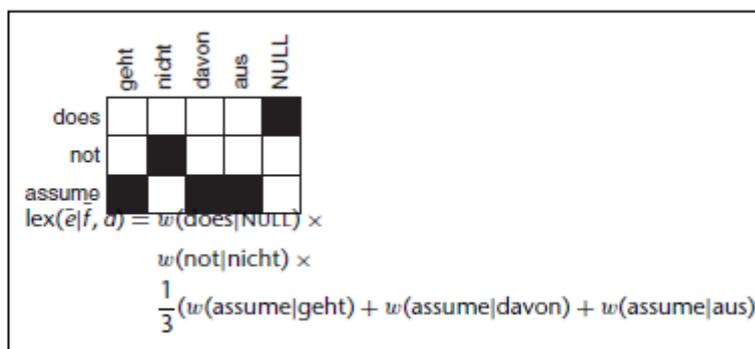


Ilustración 12: Ponderación léxica (Koehn 2010, 140)

Koehn nos da un ejemplo ilustrativo. La frase en inglés *does not assume* está alineada con la frase en alemán *geht nicht davon aus*. El peso o ponderación léxica de este par depende de tres factores: uno por cada palabra en inglés. *Not* está alineado con *nicht*, o sea que el factor es $w(\text{not}/n\text{icht})$. Mientras que *does* no está alineado con ninguna palabra en el otro idioma o sea que el factor debe transcribirse como $w(\text{does}/NULL)$. Finalmente *assume* está alineada con *geht davon aus*, por lo que el factor es la media de las tres probabilidades de traducción de las palabra correspondientes (Koehn 2010, 139–140).

2. Tecnologías de la traducción, traducción automática y sus problemas

Se han propuesto muchos métodos diferentes para la ponderación léxica, la mayoría de ellos están inspirados en los modelos de IBM basados en palabras que han demostrado ser los más eficaces.

5) Para algunos modelos es importante el número de palabras en las frases traducidas, por ello con el modelo de *Word penalty* se puede indicar si se prefieren traducciones más cortas o no. Este modelo cuenta el número de palabras en las traducciones e introduce una penalización por palabras, con la que se agrega un factor de ω para cada palabra resultante y de esta manera evitar que una traducción sea demasiado corta (o muy larga). Es decir si $\omega < 1$ se le da mejor puntuación a las traducciones más cortas. Si $\omega > 1$ se prefieren las traducciones más extensas. El incluir este parámetro puede ayudar a ajustar la traducción e incluso mejorar la calidad significativamente.

6) Con el modelo de clasificación *Phrase penalty* se miden el número de frases bilingües utilizadas para producir la traducción y se incluye —de manera análoga a la calificación por palabras— otro factor a tener en consideración, el de la longitud de las traducciones. Al emplear este modelo se podrá escoger si se prefiere tener menos frases en la traducción $\rho < 1$ (es decir, se prefieren frases más largas) o si se prefieren más frases (lo que implica que se tendrán frases más cortas) $\rho > 1$.

En la práctica, si el modelo tiene la opción de preferir una frase más larga en la traducción, tiende a usarla porque las frases más largas incluyen más contexto. Por supuesto, las frases más largas son menos frecuentes y por lo tanto estadísticamente menos fiables, pero para ello se utiliza el modelo de ponderación léxica que se encarga de descartar los pares de frases malas.

6) Esta última afirmación nos deja ver que los mejores resultados se encuentran si se combinan todos los modelos. No todos los modelos son iguales y por ello no tienen que tener el mismo valor. El objetivo del método de entrenamiento de un motor de traducción automática es averiguar cuánta importancia se debe dar a cada uno de los parámetros o modelos y ello depende del encargo de traducción. Esta configuración o entrenamiento discriminatorio se denomina *parameter tuning* o ajuste de parámetros (Koehn 2010, 266).

2. Tecnologías de la traducción, traducción automática y sus problemas

$$p(\text{source} | \text{target}) = \lambda_1 h_1(\bullet) \lambda_2 h_2(\bullet) + \dots \lambda_8 h_8(\bullet)$$

Donde λ_i es el valor que se le da al modelo h_i

1. h_1 : *p*translation model (source|target)
2. h_2 : *p*translation model (target|source)
3. h_3 : *p*language model (target)
4. h_4 : *p*reordering model (reordering operation|source, target)
5. h_5 : lex(source|target)
6. h_6 : lex(target|source)
7. h_7 : wordcount (target)
8. h_8 : phrasecount (source, target)

Para el *parameter tuning* se necesitan 1) millones de traducciones que permitan la evaluación automática de la calidad de la traducción; 2) Un corpus para la sintonización, que sea diferente del corpus de entrenamiento (≈ 3000 frases en el idioma origen con su correspondiente traducción humana). Este proceso que se desarrolla automáticamente y se inicia dándole valores aleatorios al peso de λ_i para encontrar el conjunto de valores que maximiza la calidad de la traducción. Se deben asignar a cada parámetro valores diferentes. Se traducen las frases del lenguaje origen del *tuning corpus* al lenguaje meta. Luego se comparan las traducciones con el corpus de referencia (traducción humana). Teniendo en cuenta las frases a traducir en el idioma origen se procede de la siguiente manera:

1. Se tienen en cuenta todas las traducciones posibles.
2. Se extraen muestras de cada traducción con los diferentes modelos para calcular la puntuación final.
3. Se selecciona la traducción con la puntuación más alta, según los parámetros establecidos.

$$\arg_t \max(\lambda_1 h_1(t, s) + \lambda_2 h_2(t, s) + \dots + \lambda_8 h_8(t, s))$$

Los procesos aquí descritos demandan mucho tiempo y requieren de un sistema de gran capacidad.

2. Tecnologías de la traducción, traducción automática y sus problemas

2.2.2.2 *Sistemas basados en ejemplos (EBMT)*

La base de estos sistemas es muy similar a la de los SMT porque utilizan un corpus bilingüe como base principal de conocimiento durante la traducción. Por eso no hacen parte de estos sistemas aquellas aproximaciones que utilizan el corpus en una fase inicial de entrenamiento para la construcción de un modelo —por ejemplo estadístico— de traducción (Tomás 2003, 21). Sin embargo, el enfoque de este corpus es fundamentalmente diferente: en lugar de analizar un gran corpus y con base en los datos obtenidos encontrar la traducción más probable, los sistemas basados en ejemplos comparan parte del texto origen con un corpus extenso basándose en el principio de analogía (Carl y Way 2003). El sistema EBMT identifica las partes que se pueden utilizar y las combinan de nuevo para la traducción. Finalmente, se intenta transformar los fragmentos traducidos —con base en ejemplos— en frases coherentes. Es por ello que los sistemas EBMT se asocian a menudo con sistemas de gestión de memorias de traducción (SGET). Sin embargo se trata aquí de sistemas diferentes ya en los SGET se requiere de la participación de los traductores humanos y los sistemas EBMT funcionan de forma autónoma (Somers, 2003b).

Según Cranias, Papageorgiou y Piperidis (1994) hay tres cuestiones fundamentales que definen los sistemas de traducción basados en ejemplos:

1) El establecimiento de la correspondencia entre las unidades en un texto (bi o multilingüe) en oraciones, frases o a nivel de palabras; 2) el desarrollo de un mecanismo para recuperar de la base de datos la unidad que mejor se adapte a la entrada y 3) la utilización del ejemplo de traducción recuperado para producir una propuesta de traducción de la frase original en cuestión.

Existen también sistemas basados en contexto (CBMT); el enfoque de la traducción automática basada en el contexto es relativamente nuevo y trabaja como SMT y EBMT con base en un corpus. Sin embargo —a diferencia de los enfoques anteriores— CBMT necesita un corpus más extenso en la lengua meta. La base del proceso de traducción la constituye un léxico bilingüe completo. Este incluye para cada palabra todas las posibles variantes de traducción y permite internamente traducciones alternativas. Con el fin de diferenciar entre la traducción apropiada y la que no lo es, se compara la traducción con ayuda de n-gramas con

2. Tecnologías de la traducción, traducción automática y sus problemas

el corpus meta. La variante que tiene más puntos en el corpus será la que se acepte. Las traducciones imposibles e improbables serán encontradas y eliminadas. De la misma manera se soluciona muchas ambigüedades (Carbonell et al. 2006).

2.2.2.3 *Sistemas híbridos*

Los sistemas híbridos son métodos de traducción automática que se caracteriza por el uso de múltiples enfoques de traducción automática dentro de un único sistema o motor. La motivación para el desarrollo de sistemas de traducción automática híbridos se deriva del hecho de que ninguno de los métodos descritos anteriormente ha logrado alcanzar un nivel satisfactorio de exactitud. Ya existen algunos motores de traducción automática híbridos funcionando que han tenido éxito y han mejorado la precisión de las traducciones. Además algunos sistemas de traducción automática populares emplean en la actualidad métodos híbridos. Entre ellas se encuentran PROMT, Systran y Asia Online.

Motor múltiple

Este enfoque de la híbrido traducción implica la ejecución de múltiples sistemas de traducción automática paralelamente. La propuesta final es generada a partir de la combinación de las propuestas de traducción de todos los subsistemas. Por lo general, estos sistemas utilizan subsistemas de traducción estadísticos y basados en reglas, pero otras combinaciones han sido exploradas: investigadores de la Universidad Carnegie Mellon han tenido cierto éxito combinando la traducción basada en ejemplos, subsistemas basados en el conocimiento y traducción estadística. SYSTRAN utiliza un motor híbrido con componentes basados en reglas y componentes estadísticos que aprenden de corpora mono y multilingües (SYSTRAN – Tecnologías de Traducción 2015).

Generación de reglas estadísticas

Este enfoque implica el uso de datos estadísticos que también se utiliza para generar reglas léxicas. La entrada es procesada según estas reglas, como si fuera realizada directamente por un sistema basado en reglas. De esta manera se trata de evitar la difícil y larga tarea la crear de un conjunto de integral de reglas lingüísticas mediante la extracción de dichas a partir de un corpus de entrenamiento. Este enfoque se presentan muchos más problemas que con los sistemas de SMT, ya que la exactitud de la traducción dependerá en gran medida de la simili-

2. Tecnologías de la traducción, traducción automática y sus problemas

tud del texto de entrada con el texto del corpus de entrenamiento. Como resultado de ello, esta técnica ha tenido más éxito en aplicaciones de dominio específico y tiene las mismas dificultades con la adaptación de dominio como muchos sistemas de traducción automática estadística.

Estadísticas guiadas por reglas

Este enfoque inverso al anterior utiliza las reglas para mejorar los resultados de la traducción automática por estadísticas. La técnica más común que se utiliza es preproceso de la entrada con un sistema de traducción automática basada en reglas. La salida del preprocesador basado en reglas se pasa a un sistema de traducción automática estadística, que produce la salida final. Esta técnica se utiliza para limitar la cantidad de información de un sistema estadístico tendrá que considerar, reduciendo significativamente la potencia de procesamiento requerida.

2.2.3 Sistemas de traducción automática basados en redes neuronales

Hoy en día no solo nos encontramos con los primeros tipos de sistemas de traducción automática basados en descripciones lingüísticas y reglas, y los sistemas más utilizados hasta el momento basados en corpus y estadísticas; en los últimos 3 años los sistemas neuronales de traducción automática, que también son un tipo de traducción automática estadística, han tomado mucha importancia y no solo en el ámbito académico sino ya existen aplicaciones concretas en sistemas de traducción y están siendo empleados por diferentes empresas y sistemas de traducción (Microsoft 2016). Los sistemas de traducción automática basados en redes neuronales o NMT utilizan como los SMT corpus. Es decir, se entrenan con base en una gran cantidad de textos y requieren de una gran capacidad sistemática para el entrenamiento. Como vimos anteriormente la base de datos en los sistemas basados en reglas son entradas en diccionarios específicos, que pueden ser de diferentes dominios o áreas específicas (química, medicina, automoción), y el modelo de datos son entradas de carácter gramatical, morfológico y semántico o reglas que se indican manualmente. En los sistemas estadísticos se emplean básicamente dos modelos: el de traducción y el del lenguaje: el modelo de traducción calcula la frecuencia en que se dan las traducciones de ciertas *phrasen* o secuencias de palabras en el corpus y el modelo del lenguaje valora la fluencia o legibilidad de las traducciones comparándolas con un corpus en el lenguaje meta.

2. Tecnologías de la traducción, traducción automática y sus problemas

En los sistemas neuronales o NMT no se emplea un modelo externo, sino un camino directo a través de una red neuronal. La idea detrás de una red neuronal artificial es tratar de reproducir el comportamiento biológico. Para ello tratan de imitar con fórmulas matemáticas una neurona a la que llegan señales de entrada con distintos pesos, que se suman para producir a una señal que dependerá de una función de activación especificada anteriormente.

Ya que en nuestro corpus no tenemos propuestas de ningún sistema de este tipo no entraremos en detalle. Habiendo dado un breve esbozo de los modelos y sistemas de traducción automática empleados actualmente nos dedicaremos ahora a clasificar desde el punto de vista teórico que tipos de problemas y errores se pueden presentar cuando utilizamos traducción automática.

2.2.4 Problemas y errores se presentan en la traducción automática

En esta parte hablaremos en primer lugar de los problemas y los errores lingüísticos que suelen presentarse en la traducción automática en general. Más adelante trataremos en particular de los problemas y errores que se presentan con este tipo de traducción en la combinación lingüística de nuestra tesis. Como problemas en la traducción automática consideraremos aquellos conflictos que exclusivamente o casi exclusivamente se presentan en la traducción automática de cualquier tipo de textos. También podemos hacer una diferenciación entre problemas generales y entre aquellos que se presentan más frecuentemente en un tipo particular de motor.

2.2.4.1 Tipos de problemas

Se han desarrollado diferentes categorizaciones de los tipos de problemas que se pueden presentar durante la utilización de sistemas de traducción automática. Algunas clasificaciones analizan los problemas según la etapa en que aparecen, otras según el motor que se emplea. En el capítulo siguiente nos referiremos en primer lugar a los tipos de problemas según la etapa en que se presentan, luego las diferentes categorías generales y finalmente a los problemas que se presentan dependiendo del motor empleado.

2. Tecnologías de la traducción, traducción automática y sus problemas

2.2.4.1.1 Clasificación de los problemas según la etapa en que se presentan

Existe una clasificación de los problemas de la traducción automática en tres categorías principales que se definen según las etapas en que éstos presentan: análisis, transferencia y síntesis (Arnold 2003).

Los problemas que encontramos en la etapa de análisis surgen debido a que el texto traducido puede ser interpretado (analizado) de diferentes maneras. Los problemas que encontramos en la etapa de síntesis, tiene que ver con el hecho de que un contenido o significado puede ser no solo interpretado, sino también expresado de diferentes maneras. Y los problemas que se pueden clasificar en la categoría de transferencia tienen que ver con las diferencias irreducibles en la forma en que diferentes lenguas pueden expresar el mismo contenido.

Problemas que tienen que ver con el análisis son según Arnold aquellos que se presentan cuando hay ambigüedad léxica en el texto original, cuando hay ambigüedad en la estructura y cuando no es clara la utilización de pronombres y otras expresiones anafóricas.

Tipo de ambigüedad	Ejemplos	
Ambigüedad léxica	<i>Our Jimmy has grown another foot</i>	<i>foot</i> : miembro corporal
		<i>foot</i> : unidad de medida
Ambigüedad estructural	<i>Some young girls and boys have arrived</i>	<i>young</i> califica a <i>girls</i>
		<i>young</i> califica a <i>girls</i> y <i>boys</i>
Utilización de pronombres	<i>The police refused to let the women demonstrate because they advocated violence</i>	<i>they</i> se refiere a <i>the women</i>
		<i>they</i> se refiere a <i>the police</i>

Tabla 43: Tipos de problemas que se presentan en el análisis según Arnold

La tarea de un componente de transferencia en la traducción automática es tomar el tipo de representación abstracta producida por el componente de análisis del idioma de origen (*source Interface Structure*) y producir algo que se pueden introducir en el componente de síntesis de la lengua meta (*target Interface structure*). El problema de la transferencia es que las estructuras no pueden ser iguales, porque las lenguas no asocian forma y el contenido de las mismas maneras.

2. Tecnologías de la traducción, traducción automática y sus problemas

Oración	Estructura
<i>I miss London</i>	[<i>sentence/press miss</i>], [np/sing/1st PRO], [np London]]
<i>Londres manques á moi</i>	[<i>sentence/press manquer</i>], [np Londres], [np London] [np/sing/1st PRO]]

Tabla 44: Problemas de transferencia según Arnold

En este ejemplo de transferencia se podrían establecer reglas que dijeran, por ejemplo, que *London* se debe traducir como *Londres* y el pronombre de la primera persona singular como el correspondiente en el otro idioma. El problema se presenta con la traducción del verbo y estructura gramatical debido a que no se pueden establecer reglas generales, ya que los idiomas en cuestión tienen palabras y estructuras diferentes para expresar el mismo contenido.

Los problemas de síntesis tienen que ver con la posibilidad que se tiene de expresar un contenido de diferentes maneras. El primer aspecto del problema es que en algunas ocasiones solo una de las formas de expresar el contenido es correcta; el segundo aspecto según Arnold es que se puede dar un problema cuando no hay forma obvia de la seleccionar cuál es la forma correcta de expresar el contenido. Por ejemplo, tenemos la oración *Sam saw a black cat*. Aquí tenemos un evento de visualización, donde Sam ejecuto la acción de ‘ver’, y lo visto era un gato negro y el evento ocurrió en el pasado. Este contenido se puede expresar en inglés de diferentes maneras

<i>Sam saw a cat. It was black.</i>
<i>Sam saw something black. It was a cat.</i>
<i>Sam saw a cat which was black</i>
<i>Sam was a black thing wich was a cat</i>
<i>A black cat was seen by Sam</i>
<i>Something happenend in the past. Sam saw a cat.</i>
<i>There was a black cat. Sam saw it.</i>

Tabla 45: Ejemplos de problemas en la transferencia

2. Tecnologías de la traducción, traducción automática y sus problemas

Para definir cuál de las formas es correcta solo se puede leer lo que se ha producido para ver si es claro y si el contenido es evidente para un posible lector.

A la clasificación general de Arnold añadiremos otros tipos de problemas, que si bien pueden localizarse en las categorías establecidas por el autor, vale la pena detallarlos.

2.2.4.1.2 Clasificación de los problemas según el tipo

Los problemas de orden lingüísticos dependen de los idiomas implicados. La solución de este tipo de problemas exige a la lingüística computacional una investigación fundamental, que muchas veces no puede ser realizada, ya que esto solo es posible si el sistema de traducción automática puede analizar o tener en cuenta cada regla o excepción gramatical en todas las combinaciones lingüísticas posibles. Existen ciertos tipos de problemas de orden lingüístico que se presentan en la traducción automática y que podemos generalizar; a continuación mencionaremos los más frecuentes. Siguiendo la tipología de errores del estudio empírico de Ramlow (2008, 83), que tenía como objetivo evaluar las traducciones automáticas de diferentes motores en la combinación lingüística francés-alemán, clasificaremos los problemas en cuatro grupos principales: problemas de orden morfológico, léxico, sintáctico y aquellos que tienen que ver con la referencia pronominal.

Problemas de orden morfológico

Problemas y errores de orden morfológico son, según Ramlow (2009, 240–242) aquellos que tienen que ver con el género, el número, el caso, el tiempo y el modo. Aquí algunos ejemplos:

Original	Traducción	Error
<i>Les comportements démographiques opposent traditionnellement deux France: (...)</i>	<i>Das demographische Verhalten stellt traditiongemäß zwei Frankreich gegenüber</i>	Error de número (Francia debe estar en plural)
<i>Le nombre de tués au volant a encore chuté l'an dernier (...)</i>	<i>Die Zahl von Todesopfern am Steuer hat das letzte Jahr (...) noch gefallen</i>	Error en el tiempo verbal, formación incorrecta del perfecto, verbo auxiliar erróneo

2. Tecnologías de la traducción, traducción automática y sus problemas

Original	Traducción	Error
<i>Elle vérifie également si l'incidence des décisions (...) sur l'environnement a été prise en compte.</i>	<i>Sie überprüft ebenfalls, ob die Wirkung der Entscheidungen (...) auf der Umwelt gut Berücksichtigungen gewesen ist</i>	Error en el caso, la preposición <i>auf</i> puede exigir en alemán el caso dativo o el acusativo, en este ejemplo se escoge dativo aunque debería ser acusativo

Tabla 46: Ejemplos de problemas de orden morfológico

También bajo este tipo de problemas debemos incluir ciertos casos de ambigüedad morfológica que pueden darse en la flexión, en la derivación o en la construcción de palabras compuestas. En la flexión nos encontramos con casos de ambigüedad morfológica si p.ej. una forma verbal en infinitivo es idéntica a otras personas y tiempos verbales y por ello puede ser interpretada en la traducción automática erróneamente. En ciertas combinaciones lingüísticas las desinencias de la conjugación verbal pueden ser iguales a pesar de que la persona o el número sean diferentes. Este es el caso de muchos verbos alemanes cuya flexión en la primera y la tercera plural es igual y además no se diferencian del verbo en infinitivo. Por otra parte la forma de cortesía *Sie* se conjuga de la misma manera que la 3ª. persona plural. Si no se tiene una referencia concreta del contexto es posible que el texto meta no corresponda con el sentido del texto origen.

En el caso de la derivación también pueden presentarse ambigüedad, ya que existen similitudes entre las formas singulares o plurales de adjetivos o sustantivos lo que dificulta la traducción automática.

Forma	Alemán	Castellano
1ª. persona plural	<i>wir waren alt</i>	nosotros éramos viejos
3ª. persona plural femenina	<i>Sie war alt</i>	Ella era vieja
3ª. persona plural masculino	<i>Er war alt</i>	El era viejo
Forma de cortesía	<i>Sie sind alt!</i>	¡(Usted) es viejo!

Tabla 47: Diferencias en el número y género

2. Tecnologías de la traducción, traducción automática y sus problemas

Problemas en la traducción de nombres compuestos son también problemas de orden morfológico. En idiomas como el alemán o el japonés, que tienden a crear nuevas palabras a través de composiciones, la relación entre los sustantivos correspondientes está prácticamente que implícita. En otros idiomas se debe hacer explícita, por ejemplo a través del uso de preposiciones. Ya que las preposiciones se utilizan de manera diferente en cada idioma, es la traducción automática de nombres compuestos un reto para los motores de traducción automática. En alemán encontramos frecuentemente palabras compuestas como el sustantivo *Dampfschiffahrtsgesellschaft* (Hutchins y Somers 1992, 85) que en castellano significa ‘compañía naviera que opera barcos de vapor’. También nos encontramos con otro tipo de composiciones, no tan complejas, que sin embargo pueden generar problemas en la traducción (Ramlow 2008, 58). En ciertos casos se puede dar una interpretación correcta con ayuda de un análisis semántico, como en el caso del sustantivo *Beileid*, en el que se puede deducir que solo una de las posibilidades tiene sentido; pero en otros casos las diferentes posibilidades de segmentación son interpretaciones semánticamente posibles.

Palabra compuesta	Segmentación	Segmentación
<i>Beileid</i>	<i>Bei+Leid</i>	<i>Beil+Eid</i>
<i>Kulturgeschichte</i>	<i>Kultur + Geschichte</i>	<i>Kult+Urgeschichte</i>
<i>Wachtraum</i>	<i>Wach+traum</i>	<i>Wacht+raum</i>

Tabla 48: Ejemplos de problemas en la traducción de sustantivos compuestos (Hutchins y Somers 1992, 85)

La transferencia de los casos puede ser problemática para la traducción automática. Hay que distinguir en qué sentido se habla de casos; si hablamos de un caso en sentido abstracto o si hablamos del caso morfológico. En sentido abstracto no podemos visualizar la asignación del caso, en el sentido morfológico, el caso si se puede ver claramente, pues hay un cambio morfológico. En sentido abstracto incluso el idioma inglés tiene casos, que a nivel morfológico solo se pueden ver en los pronombres: *I* o *he* (nominativo), *me* o *him* (acusativo), *my* o *his* (genitivo).

2. Tecnologías de la traducción, traducción automática y sus problemas

Problemas lingüísticos de orden léxico

Ramlow también se refiere en su estudio a problemas de orden léxico y los divide a su vez en seis tipos: aquellos que tienen que ver con traducción de homógrafos (aquí incluye también la polisemia), problemas terminológicos, problemas en la traducción de preposiciones, problemas que se dan en la traducción de *phrasen* como colocaciones, errores léxicos que se dan porque no se traducen palabras y finalmente deja un último grupo para incluir aquellos problemas y errores léxicos que puedan aparecer en su estudio empírico y no se ajusten a ninguna de las categorías mencionadas.

a) Homógrafos, ambigüedad léxica y polisemia

Como ya habíamos mencionado en el capítulo sobre los “Problemas léxico-semánticos” en la página 54, los homógrafos y la polisemia constituyen un problema para la traducción en general. Los homógrafos pueden ocasionar problemas especialmente en la traducción automática, donde no se puede identificar sin el contexto, cuál es el significado correspondiente en el caso particular. Si los homógrafos son de categorías diferentes se puede resolver la ambigüedad en el análisis sintáctico, pero no si pertenecen a la misma categoría. En la oración *Gas pump prices rose last time oil stocks fell* encontramos algunas palabras que tiene al menos una ambigüedad en dos sentidos: puede ser sustantivos o verbos y *last* incluso puede ser sustantivo, verbo, adjetivo o adverbio. A pesar de las ambigüedades, solo hay una forma correcta de analizar de esta frase, aunque requiere algo más que el simple conocimiento de las posibles combinaciones de categorías para llegar en ello. En estos casos la traducción automática tiene que recurrir a la información semántica y asignar rasgos semánticos específicos que son característicos en determinadas construcciones sintácticas (Hutchins y Somers 1992, 87).

Original	Traducción
<i>La création du Fonds de Cohésion</i>	<i>Die Schöpfung des Kohäsionsfonds</i>
<i>...par rapport aus objectifs fixés</i>	<i>... im Vergleich zu den befestigen Zielen</i>

Tabla 49: Problemas en la traducción automática de homógrafos

En el ejemplo anterior de Ramlow (2008, 84) podemos ver como no siempre se escoge el significado correcto en los casos de homografía. En francés como en castellano utilizamos

2. Tecnologías de la traducción, traducción automática y sus problemas

el sustantivo ‘creación’ tanto para denominar la acción o el efecto de crear, como el acto de creación divina o incluso para denominar una obra de ingenio, de arte o artesanía muy laboriosa; en alemán existen sustantivos diferentes para estas acciones, y por ello la traducción escogida no es la correcta. *Schöpfung* tiene el sentido de creación divina o artística, mientras que los sustantivos *Gründung* o *Errichtung* se ajustan más al contexto. En el segundo ejemplo (Ramlow 2009, 185) vemos como la traducción del adjetivo *fixés* si bien es correcta no es la adecuada para el contexto.

La resolución de ambigüedades fue considerada ya por Hutchins y Somers como uno de los problemas lingüísticos más importantes con los que se enfrenta la traducción automática (Hutchins y Somers 1992, 95). En otro ejemplo interesante de estos autores *Foot heads arms body* podemos ver que en algunos casos solo se puede resolver la ambigüedad si se tiene un ‘conocimiento del mundo real’: *Foot*, es un político británico, que fue nombrado presidente de una comisión que investigaba el control de armas. Según Hutchins y Somers existen tres tipos fundamentales de ambigüedades léxicas: las ambigüedades categoriales, homógrafos o polisemias y la ambigüedad de transferencia o traducción (1992, 85).

La ambigüedad categorial se da cuando a una palabra se le puede asignar más de una categoría gramatical o sintáctica dependiendo del contexto. Por ejemplo *light* en inglés puede ser un sustantivo, un verbo o un adjetivo. Algunas de estas ambigüedades se pueden resolver si se tiene en cuenta la flexión morfológica: *rounding* es sin duda una forma verbal, mientras que *rounds* puede ser un nombre en plural o la tercera persona del verbo en presente. Pero se recurre más frecuentemente al análisis sintáctico para resolver este tipo de ambigüedades. En la oración *The cowboy started to round up the cattle* el contexto sintáctico determina que *round* solo puede ser un verbo (1992, 85–86).

La polisemia se da por una ambigüedad sistemática debido a la relación etimológica o debido a una relación de los significados individuales que tienen un significado básico común o significado común central que —en muchos casos— se debe a la presencia de una raíz etimológica común.

Los motores de traducción automática no pueden distinguir siempre en qué casos se está usando una palabra o frase en un sentido diferente, por ejemplo, al más usado o al grama-

2. Tecnologías de la traducción, traducción automática y sus problemas

ticamente más indicado. Mientras que las ambigüedades categoriales, la homografía y la polisemia son ambigüedades léxicas que causan problemas al analizar el texto de origen, en la traducción nos encontramos con otro tipo de ambigüedades, las ambigüedades de transferencia o de traducción que se dan cuando una palabra en la lengua de origen puede traducirse de distintas maneras. Este tipo de ambigüedades causan muchos problemas en la traducción automática.

A estos los problemas léxicos mencionados tenemos que añadir las diferencias léxicas estructurales. Una lengua puede tener un vacío léxico, cuando hay una palabra o frase corta, cuyo significado no se puede expresar en el idioma meta sin una nota explicativa; también cuando a un concepto representado por una palabra en el idioma origen le corresponde varios conceptos y palabras en el idioma meta. Este es el caso de la palabra *leg* en inglés la que en castellano o francés se puede traducir de diferentes maneras dependiendo del caso como vemos en el ejemplo original de Hutchins & Somers (1992, 101) utilizado también por Jurafsky & Martin (2006, 806).

Inglés	Francés	Castellano
<i>leg</i>	<i>jambe</i> (ser humano)	pierna (ser humano)
	<i>patte</i> (animal)	pata (animal)
	<i>piéd</i> (silla o mesa)	pie (silla)
	<i>étape</i> (de un día)	etapa (día)

Tabla 50: Ejemplo de diferencia léxica entre el inglés, francés y castellano

b) Problemas léxicos de orden terminológico

Como su nombre lo indica tiene que ver con las dificultades o errores que se pueden presentar al no utilizarse correctamente la terminología especial del área de conocimiento o especialización. También hay que tener en cuenta que algunos errores de terminología pueden presentarse debido a la homografía. A continuación algunos ejemplos de errores terminológicos analizados por Ramlow (2008, 84).

2. Tecnologías de la traducción, traducción automática y sus problemas

Original	Traducción automática
<i>Organisation mondiale du commerce</i>	<i>Weltorganisation des Handels</i>
<i>On opérè en droit la distinction suivante: les personnes physiques, (...) les personnes morales (sociétés, associations)</i>	<i>Man operiert die folgende Auszeichnung in Recht: die korperliche Personen (...), die moralische Personen (Gesellschaften, Vereine)</i>

Tabla 51: Problemas léxicos de orden terminológico

En estos casos la traducción automática no reconoce la terminología jurídica y se traduce palabra por palabra, por ello no se traduce correctamente el nombre propio Organización Mundial del Comercio como *Welthandelsorganisation* (Ramlow 2009, 185). En el segundo ejemplo (2008, 106) solo 2 de los seis motores de traducción automática analizados por el autor traducen correctamente el término jurídico de personas naturales y jurídicas a su equivalente en alemán (*natürliche - juristischen Personen*). Los problemas y errores de orden léxico que tienen que ver con la terminología pueden evitarse si se usa el diccionario terminológico correspondiente en el proceso TA.

c) Preposiciones

Bajo esta categoría incluye el autor aquellos casos en que la traducción de las preposiciones es errónea, es decir se emplea la preposición incorrecta, pero no se cambia la estructura sintáctica de la oración.

Original	Traducción automática
<i>pour répondre à une infraction</i>	<i>..um zu antworten an einen Verstoß..</i>

Tabla 52: Problemas léxicos en la traducción automática de preposiciones (Ramlow 2009, 185)

Si bien en este ejemplo se traduce correctamente el orden sintáctico de la oración del francés al alemán, la preposición utilizada en alemán es *auf* y no *an*.

d) Phrasen

Todos los problemas y errores que pueden presentarse al traducir colocaciones, *phrasen*, giros y construcciones léxicas de un idioma a otro se agrupan en esta categoría. Veamos un ejemplo del autor:

2. Tecnologías de la traducción, traducción automática y sus problemas

Original	Traducción automática
<i>Autant d'appréciations qui laissent de marbre le syn- dicats d'enseignants</i>	<i>So viele Einschätzungen, die aus Marmor die Lehrgewerkschaften lassen</i>

Tabla 53: Problemas léxicos en la traducción automática de *phrasen*

En el ejemplo (Ramlow 2009, 185) podemos observar como la expresión *laisser de marbre*, se trata de traducir palabra por palabra y no con su correspondiente en alemán *kalt lassen*. Pero no solo en la combinación lingüística del ejemplo, en todos los idiomas existen construcciones gramaticales complejas, como las combinaciones comparativas como en inglés “*all/none of*”, en alemán “*je (adverbio)/desto (adverbio)*” o en castellano “cuanto más/ más”. Ramlow afirma que a pesar de su complejidad la traducción automática de colocaciones puede realizarse sin problemas, si estas son identificadas como tales. En el siguiente ejemplo de Ramlow (2008, 62) podemos ver que la traducción de ciertas frases idiomáticas puede ser problemática cuando estas pueden ser interpretadas como tales, pero también tener un sentido normal:

Original	Sentido	Traducción
<i>Er hat den Löffel abgegeben</i>	Traducción palabra por palabra	El entregó la cuchara
	Frase idiomática	El murió

Tabla 54: Ejemplo de Ramlow de interpretaciones posibles en una frase idiomática

Al producir textos no solo juega un papel importante la selección libre de palabras sino sobre todo el principio idiomático. Según este principio no se escogen las palabras de manera aislada sino en grupos o incluso en frases, que se encuentran a disposición del usuario/hablante del idioma. Colocaciones son construcciones de dos o más palabras, que a pesar de componerse de varias palabras o morfemas son escogidas conjuntamente en una sola acción (Caro Cedillo 2004, 15–17). Es decir son palabras que aparecen juntas en una determinada relación. No pertenecen a las fraseologías o expresiones idiomáticas porque su significación se puede deducir de la suma de los significados de los componentes. Existen colocaciones que usamos diariamente en la comunicación como ‘poner la mesa’ y otras que pertenecen

2. Tecnologías de la traducción, traducción automática y sus problemas

a una temática concreta como el lenguaje jurídico (dictar un auto de procesamiento) o el de las finanzas (girar una letra de cambio).

A pesar de que su significación puede entenderse las colocaciones pueden ser problemáticas en la traducción automática ya que en muchos casos los componentes solo se utilizan en un sentido particular y no pueden ser reemplazados por sus sinónimos. También pueden ser problemáticas cuando en idiomas como el alemán o el castellano pueden variar dependiendo de la región en la que se apliquen. Mientras que en España se utiliza ‘alquilar un coche’ en Chile por ejemplo se emplea ‘rentar un coche’ (Caro Cedillo 2004, 40–41).

Traducción correcta	Original	Traducción automática (Google)
Diversificar la cartera	<i>Mischung des Portefeuilles</i>	Mezclar la cartera

Tabla 55: Problemas que pueden aparecer en la traducción automática de frases idiomáticas

Hablamos de frases idiomáticas o modismos como un grupo de palabras, con varias partes, que forma una unidad semántica que no se puede derivar de las partes individuales. Las modificaciones o sustitución de los elementos son prácticamente imposibles. La traducción automática de este tipo de frases idiomáticas puede ser errónea especialmente si se lleva a cabo con un motor basado en reglas y no en estadísticas, ya que la traducción correcta no será la traducción exacta de la frase, sino lo que esta implica o quiere decir.

Expresiones idiomáticas son, según Arnold, “expressions whose meaning cannot be completely understood from the meanings of the component parts” (Arnold et al. 1994, 114). Algunas expresiones idiomáticas tienen un equivalente similar en otros idiomas, más no siempre se puede encontrar el equivalente exacto.

Expresión en inglés	Equivalente en alemán	Equivalente en castellano
<i>If Sam kicks the bucket, her children will be rich (Arnold et al. 1994:115)</i>	<i>Wenn Sam den Löffel abgibt, seine Kinder werden reich</i>	Si Sam estira la pata , sus hijos serán ricos

Tabla 56: Expresiones idiomáticas correspondientes en alemán y castellano

2. Tecnologías de la traducción, traducción automática y sus problemas

A menos que se entrene con anterioridad el sistema de tal manera que la traducción de la expresión *kick the bucket* tenga el mismo significado que el verbo ‘morir’, la traducción automática no podrá dar una propuesta correcta. Además si se quiere reproducir la intención comunicativa del texto origen, no solo debería reproducirse el sentido sino además encontrarse una expresión idiomática equivalente en el idioma meta.

También las metáforas y las metonimias son construcciones que pueden ser problemáticas para la traducción automática, porque para su correcta interpretación se requiere de un conocimiento profundo tanto del idioma origen como del idioma meta. La resolución de anáforas es también un problema de orden lingüístico con el que se enfrenta la traducción automática. Anáfora es una referencia indirecta a una entidad, generalmente un sustantivo, mencionada de forma explícita con anterioridad en el texto, como anáforas suelen utilizarse frecuentemente pronombres personales y demostrativos o expresiones como ‘el último’ o ‘lo anterior’. Para poder resolver la anáfora se debe identificar el sustantivo a que se refiere, el denominado antecedente. Esta tarea no es siempre fácil, como se ilustra en el ejemplo:

<i>The monkey ate the banana because it was hungry.</i>
- <i>The monkey ate the banana because it was ripe.</i>
- <i>The monkey ate the banana because it was tea-time</i>

Tabla 57: Anáforas

En todos estos casos el pronombre ‘it’ se refiere a algo diferente, aquí se deben emplear otro tipo de estrategias para resolver la ambigüedad, por ejemplo restricciones de comparación: *be hungry* exige un sujeto animado y es más probable que *be ripe* se refiera a una fruta que a un animal (Hutchins y Somers 1992, 95).

Problemáticas para la traducción automática son las anáforas, cuando se traduce a idiomas donde cada opción tendría una traducción diferente:

Original	Traducción al francés
<i>The soldiers shot at the women and some of them fell</i>	<i>Les soldats ont tiré sur les femmes et quelques-unes sont tombées</i>

2. Tecnologías de la traducción, traducción automática y sus problemas

Original	Traducción al francés
<i>The soldiers shot at the women and some of them missed</i>	<i>Les soldats ont tiré sur les femmes et quelques-uns ont raté</i>

Tabla 58: Diferencia de tratamiento de anáforas en inglés y francés

En el ejemplo de Hutchins y Somers podemos ver que para poder traducir a ciertos idiomas como el francés, el castellano o el alemán es necesario saber si *some of them* se refiere a las mujeres o a los soldados, no solo para precisar el género del pronombre sino también para hacer concordar el participio pasado (*tombé(e)s*). Según los autores para poder traducir correctamente no solo se requiere de un conocimiento lingüístico sino también de un conocimiento contextual o del mundo real (1992: 96).

Los complementos restrictivos que analiza Alcaraz Varó (2007, 20) en su artículo sobre los problemas metodológicos en la traducción del inglés jurídico, son unidades lingüísticas (palabras, sintagmas u oraciones) que limitan el significado de otra unidad. El autor expone cómo el alcance de estos complementos dificulta la interpretación. En castellano el sintagma “El supuesto instigador y cabecilla de la revuelta” puede tener dos interpretaciones, dependiendo de si el adjetivo “supuesto” califica solo al sustantivo “instigador” o si califica también al “cabecilla”. Es decir que nos entramos aquí con un caso particular de ambigüedad estructural. La interpretación del alcance de tales complementos ha sido motivo no solo de confusión, sino también de múltiples controversias en los tribunales fiscales estadounidenses que han llevado a los jueces incluso a evitar tales restricciones sintácticas para no caer en el absurdo. El autor nos da varios ejemplos como el caso jurídico de la interpretación de *in the vicinity of a prohibited place*, donde un acusado afirmó que no encontrarse en la vecindad del lugar prohibido (*in the vicinity*), sino en el lugar prohibido (*in the prohibited place*), por lo cual no había cometido ningún delito. Casos como este, y no nos referimos solo a aquellos que como en el ejemplo pueden llevarse al absurdo, se presentan realmente en la traducción y son especialmente problemáticos en la traducción automática.

También nos encontramos con ambigüedad en el alcance de los cuantificadores; este tipo de ambigüedad se produce cuando no se puede precisar con certeza qué tanto implica un

2. Tecnologías de la traducción, traducción automática y sus problemas

cuantificador como ‘alguno/as’, ‘todos/as’ o ‘ninguno/a’. Vemos uno de los ejemplos que utilizan Hutchins y Somers (1992, 96) para ilustrar la problemática.

Texto original	Interpretaciones posibles
All women don't like fur coats	A todas las mujeres no les gustan los abrigos de piel.
	No a todas las mujeres les gustan los abrigos de piel, solo a algunas.
	No hay ninguna mujer a la que le gusten los abrigos de piel.

Tabla 59: Ambigüedad por ampliación de cuantificadores

e) Falta de traducción

Aquí agrupa Ramlow los errores que se presentan cuando no se traduce una o varias palabras. En algunos casos, los menos graves, se dejan en el idioma original; pero también puede suceder que se ignoren por completo dejando un vacío léxico. También puede ocurrir que este tipo de omisiones solo se puedan detectar si se hace una comparación directa del segmento original y la traducción.

Original	Traducción automática
... que cela est possible (...) pour les très petits nombres das das (...) für die sehr kleinen nombres möglich ist ⁵ .
(...) les différentes pratiques de calcul: mental, posé, approché, exact et instrumenté .	(...) verschiedene Rechenpraktische: geistig, ruhig, näher gerückt, genau und.

Tabla 60: Problemas por la omisión o falta de traducción de palabras

f) Otro tipo de problemas de orden léxico

Como mencionamos anteriormente Ramlow deja una categoría abierta para incluir en ella los problemas o errores que no se ajusten a las categorías mencionadas anteriormente. Allí agrupé aquellas traducciones que si bien no eran erróneas en sentido estricto, eran idiomáticamente incorrectas y por ello deberían en todo caso ser corregidas, en caso tal que la traducción quiera ser por ejemplo publicada.

⁵ (Ramlow 2009, 185)

2. Tecnologías de la traducción, traducción automática y sus problemas

Original	Traducción automática
... ils seront très prochainement analysés.	... werden sie sehr demnächst analysiert werden.

Tabla 61: Otros problemas de orden léxico

En el ejemplo (Ramlow 2009, 238–239) se traduce el adverbio *prochainement* con el adverbio *demnächst*, pero en alemán es más común la utilización en estos casos del adverbio *auf* o la expresión *in Kürze*.

2.2.4.1.2.1 Problemas de orden sintáctico

Este tipo de problemas pueden darse en cinco niveles: a nivel del ordenamiento de las palabras, a nivel estructural, en el uso de preposiciones o cuando la función de palabras o grupos de palabras no se reconoce correctamente, de tal manera que se pierde la referencia interna de estas. Ramlow también deja aquí un último grupo en el que se incluirán los problemas y errores de tipo sintáctico que no se ajusten a los enumerados anteriormente.

a) Ordenamiento de palabras o grupos de palabras

Veamos a continuación un ejemplo de un problema de orden sintáctico que se da por el posicionamiento erróneo de las palabras (Ramlow 2009, 242–243):

Original	Traducción automática
<i>Elle met davantage l'accent sur l'inflation</i>	<i>Sie legt den Akzent auf die Inflation mehr</i>
	<i>Sie tut mehr den Akzent auf die Inflation</i>

Tabla 62: Problemas en el ordenamiento de las palabras

En el ejemplo anterior el adverbio *mehr* fue colocado por dos sistemas de traducción automática en la posición errada, correcto sería *Sie legt **mehr** den Akzent auf die Inflation*. Este tipo de problema no se presenta solo a nivel de palabras, también en grupos de palabras o colocaciones:

Original	Traducción automática
<i>L'OMC compte à présent 144 membres, dont la Chi-</i>	<i>OMC rechnet jetzt 144 Mitglieder, unter denen China</i>

2. Tecnologías de la traducción, traducción automática y sus problemas

Original	Traducción automática
<i>ne et Taiwan, qui ont adhéré en 2001.</i>	<i>und Taiwan, die gehaften haben 2001.</i>
	<i>Die World Trade Organisation zählt jetzt 144 Glieder das, China, und Taiwan, die gehaften haben im Jahre 2001.</i>

Tabla 63: Problemas en la ordenación de oraciones relativas

En ambas propuestas de traducción automática podemos ver que hay problemas en el ordenamiento de las palabras que hacen parte de la oración relativa *qui ont adhéré en 2001*. En alemán debe colocarse el verbo conjugado al final de la oración, es decir, correcto sería *die im Jahre 2001 gehaften haben*.

b) Estructura sintáctica

Cada idioma tiene una estructura sintáctica que puede ser en mayor o menor grado diferente en comparación con otro idioma. Es por ello que suele presentarse errores de traducción cuando se transfiere la estructura de un idioma al otro, sin tener en cuenta las reglas y normas sintácticas correspondientes (Ramlow 2009, 244):

Original	Traducción automática	Traducción correcta
<i>Ansi, à la fin des années 1980, (...)</i>	<i>So, am Ende der Jahre 1980 (...)</i>	<i>Am Ende der 1980er Jahre, (...)</i>
<i>Membres : 144 pays en 2002, représentant plus 90 % du commerce mondial.</i>	<i>Mitglieder: 144 Länder 202, mehr als 90 % des Welthandels darstellend</i>	<i>144 Mitglieder, die mehr als 90 % des Welthandels darstellen.</i>

Tabla 64: Problemas debidos a las diferencias en la estructura sintáctica

c) Preposiciones

La traducción de preposiciones no solo es problemática a nivel léxico, sino también sintáctico. Pude darse el caso de que la preposición falte en la traducción, porque en el idioma original no es necesaria, o también puede darse el caso que se utilice una preposición a pensar de que en el idioma meta no es necesario (Ramlow 2009, 187,246).

2. Tecnologías de la traducción, traducción automática y sus problemas

Original	Traducción automática	Traducción correcta
<i>Toutes les banques centrales de l'UE font partie de SEBC.</i>	<i>Alle Zentralbanken von der EU gehören zu ESZB.</i>	<i>Alle Zentralbanken der EU gehören zu ESZB.</i>
<i>Ce vieillissement démographique (...) est lié à plusieurs facteurs:</i>	<i>Dieses demoographische Altern (...) ist mehreren Faktoren verbunden:</i>	<i>Dieses demoographische Altern (...) ist mit mehreren Faktoren verbunden:</i>

Tabla 65: Problemas sintácticos en la traducción de preposiciones

Uno de los errores gramaticales más frecuentes es la utilización incorrecta de las preposiciones o sintagmas preposicionales, ya que se tiende o bien a transcribirlas en el idioma meta como si fueran equivalentes, o simplemente porque el uso de las preposiciones es diferente en los idiomas implicados:

Original	Traducción
Encontró un libro sobre la mesa	<i>He found the book on the table</i>
Encontró un libro sobre la guerra civil	<i>He found the book about/on</i>

Tabla 66: Ejemplo del uso diferente de las preposiciones en castellano e inglés

d) Referencia sintáctica

Problemas a este nivel se dan cuando la función sintáctica de palabras o grupos de palabras no se reconoce y no se traduce correctamente su relación:

Original	Traducción automática
<i>Le phénomène ne puet que s'accélérer au cours des prochaines décennies, en raison du vieillissement der nombreuses gébérations (...)</i>	<i>Das Phänomen kann nur im Laufe der nächsten Jahrzehnte, auf Grund des Alterns zahlreicher Generationen schneller werden, (...)</i>
	<i>Das Phänomen kann während der nächsten Jahrzehnte nur aufgrund des Alterns der zahlreicher (...) Generationen schneller werden.</i>

Tabla 67: Problemas en la referencia sintáctica

2. Tecnologías de la traducción, traducción automática y sus problemas

La expresión en francés *en (...) que (...)* se refiere a *s'accélérer*, en la primera traducción al alemán se relaciona con *im Laufe der nächsten Jahrzehnte*, en la segunda propuesta de traducción con *aufgrund des Alterns*. En ambos casos la referencia es incorrecta. La traducción correcta debería posicionar el adverbio *nur* frente a *schneller werden* (Ramlow 2009, 247–248).

Si bien Ramlow no habla directamente del tema incluiremos en este tipo de problemas la ambigüedad en las funciones sintácticas según la definición de Hutchins y Somers. Ambigüedades en las funciones sintácticas se presentan, como su nombre lo indica, cuando no hay claridad en la función de alguno de los componentes. En el ejemplo a continuación (Hutchins y Somers 1992, 88–89) podemos ver cómo el complemento “con el telescopio” puede tener dos funciones diferentes y así interpretarse de dos maneras distintas.

Original	Posibles interpretaciones
<i>The man saw the girl with the telescope</i>	a) El hombre vio a la niña con la ayuda de un telescopio.
	b) El hombre vio a la niña que tenía un telescopio.

Tabla 68: Ejemplo de ambigüedad estructural

En algunos casos se puede distinguir fácilmente qué significado tiene una palabra en una frase en un determinado contexto; es decir, se está en capacidad de distinguir los diferentes significados o interpretaciones que puede tener una palabra en una frase y optar, según el contexto, por una de las posibles alternativas. En algunas oraciones complejas, el análisis sintáctico puede arrojar varias posibilidades. Para aclarar el tipo de ambigüedad daremos algunos ejemplos. Se puede dar una ambigüedad en construcciones con la forma nominalización + referencia + sustantivo:

Original	Posibles interpretaciones y traducciones
<i>Das Feiern der Mannschaft ...</i>	a) el equipo tuvo una celebración.
	b) el equipo fue celebrado, por ejemplo por sus seguidores.

Tabla 69: Ambigüedad sintáctica en determinadas construcciones

2. Tecnologías de la traducción, traducción automática y sus problemas

También Dorr (1994) se refiere a los problemas de orden sintáctico y enumera ciertos tipos de divergencias que se presentan cuando dos idiomas utilizan la misma construcción para diferentes propósitos o utilizan diferentes construcciones con el mismo propósito. Según este autor podemos encontrarnos con siete tipos diferentes de divergencias:

Tipo	Idioma 1	Idioma 2
<i>Thematic</i>	<i>I like Mary</i> <i>Mary</i> = es aquí un objeto verbal	A mí me gusta María María = es el sujeto
<i>Promotional</i>	<i>John usually goes home</i> El modificante es ' <i>usually</i> ' (adverbial)	Juan suele ir a casa El modificante es el verbo ' <i>soler</i> '
<i>Demotional</i>	<i>I like to eat</i> ' <i>like</i> ' es el verbo principal (<i>to eat</i> es un argumento de <i>like</i>)	<i>Ich esse gern</i> ' <i>gern</i> ' adverbio y solo un adjunto <i>essen</i> es el verbo principal
<i>Structural</i>	<i>John entered the house</i> ' <i>the house</i> ' es como objeto PN	Juan entró en la casa mientras que ' <i>en la casa</i> ' es un PP
<i>Conflational</i>	<i>I <u>stabbed</u> John</i> Una sola palabra (<i>stab</i>) describe toda la acción	Yo le <i>di puñaladas</i> a Juan Se necesitan más palabras para describir la acción (dar puñaladas)
<i>Categorial</i>	<i>I am hungry</i> El predicado es un adjetivo	<i>Ich habe Hunger</i> El predicado es nominal
<i>Lexical</i>	<i>John <u>broke</u> into the room</i> El suceso se realiza con el verbo principal	Juan forzó la entrada al cuarto Se necesita de un verbo diferente

Tabla 70: Tipos de divergencias según Dorr

Las divergencias se encuentran en la comparación directa entre dos idiomas, donde nos podemos dar cuenta que los idiomas implicados son no solo léxica sino sintácticamente diferentes. Es básicamente por esto que no se puede dar una traducción palabra por palabra que sea correcta, por ejemplo cuando un concepto en la lengua origen puede tener más de una posibilidad de traducción en el lenguaje meta.

2. Tecnologías de la traducción, traducción automática y sus problemas

Alemán	Inglés
<i>Mein Zug fährt gerade ab</i>	<i>My train is leaving</i>

Tabla 71: Ejemplo de diferencias estructurales

En el ejemplo anterior el adverbio alemán *gerade* se traduce al inglés con el aspecto progresivo del verbo *is leaving*. Otro tipo de diferencias estructurales las encontramos por ejemplo en la traducción de idiomas eslavos en los que no hay artículos, lo que dificulta la traducción a idiomas que si los tienen.

e) Otros tipos de problemas

Ramlow encontró en su estudio casos problemáticos en la traducción de comparativos o superlativos. En el primer ejemplo que vemos en la tabla si se traducen palabra por palabra la traducción no se ajusta a la estructura sintáctica del alemán (Ramlow 2009, 188). También la omisión o uso inadecuado de artículos en la traducción es un problema que el autor clasifico bajo este tipo, podemos ver en el segundo ejemplo, que en alemán se requiere del artículo en genitivo para darle sentido a la oración (2009, 251).

Original	Traducción automática	Traducción correcta
<i>Par une fécondité nettement plus élevée...</i>	<i>... durch eine deutlich mehr hohe (...) Fruchtbarkeit...</i>	<i>... durch eine deutlich höhere Fruchtbarkeit...</i>
<i>L'Organisation mondiale du commerce se trouve au coeur du système de régulation du commerce mondial.</i>	<i>Die Weltorganisation des Handelst findet sich am Herzen des Systems der Regelung Welthandel.</i>	<i>Die Welthandelsorganisation findet sich am Herzen des Systems der Regelung des Welthandels.</i>

Tabla 72: Otro tipo de problemas de orden sintáctico

A continuación nos referiremos a la otra categoría de problemas que Ramlow encuentra en su estudio.

Problemas y errores en la pronominalización

Ramlow se refiere a los problemas o errores en la pronominalización, es decir, a la dificultad que se presenta en la traducción automática al tratar de adaptar los pronombres a la

2. Tecnologías de la traducción, traducción automática y sus problemas

gramática del idioma meta; para ilustrar a lo que se refiere el autor veamos el siguiente ejemplo. (2009, 250–253).

Original	Traducción automática
... la densité moyenne de 11 habitants au km ² est nettement plus faible que <i>celle</i> de ser voisins européens (...)	Die durchschnittliche Dichte von 111 Einwohner in Km2 ist deutlich schwächer als <i>dieses</i> seiner europäischen Nachbarn (...)

Tabla 73: Problemas de pronominalización

El pronombre demostrativo femenino *celle* se refiere a *densité moyenne*, cuya traducción en alemán es *durchschnittliche Dichte*. *Dichte* es un sustantivo femenino y por ello requiere del pronombre *die* o *diejenige* como traducción correspondiente al pronombre *celle* pero en la traducción se utiliza el pronombre demostrativo neutro *dieses*.

2.2.4.1.2.2 Problemas o errores formales

Existen otro tipo de problemas que son importantes especialmente en la traducción automática, se trata de los problemas de tipo formal, como aquellos que tienen que ver con la puntuación. Es evidente que se pueden presentar cuando se asume la puntuación del idioma origen y no se ajusta a las reglas del idioma meta. A continuación podemos ver como en la propuesta de traducción se coloca una coma en el mismo lugar que se encuentra en el texto original, a pesar de que en este caso no es necesaria.

Original	Traducción automática
Le nombre de tués au volant a enconre chuté l'an dernier, à 4703	Die Zahl von Todesopfer am Steuer hat das lezte Jahr, in 4803 noch gefallen
Chandque année, la Commission et les États membres vérifient si le pacte est respecté.	Jedes Jahr, die Commnission und die Mitgliedstaaten überprüfen, ob der Pakt eingehalten wird.

Tabla 74: Errores de puntuación

Las reglas de puntuación, como en los ejemplos anteriores (Ramlow 2009, 257) pueden variar dependiendo del idioma. También el uso de mayúsculas y minúsculas puede ser

2. Tecnologías de la traducción, traducción automática y sus problemas

diferente y también problemático en la traducción automática, si se reproduce lo que se da en el original.

2.2.4.1.2.3 Problemas en el texto origen

Mientras que en la traducción humana el traductor puede detectar errores cometidos en el texto origen y si estos no son muy graves hacer una traducción correcta, en la traducción automática cualquier tipo de error en el texto origen se reproducirá en el texto meta o no permitirá que se produzca un texto meta correcto.

No solo podemos encontrarnos con errores ortotipográficos en el original. Estos errores no son tan problemáticos ya que algunos motores de traducción automática tienen capacidad para corregirlos. Pero los errores gramaticales dificultan enormemente el proceso de traducción automática. Solo un poseedor puede detectarlos y corregirlos satisfactoriamente.

Errores de estilo pueden ocasionar menos problemas, sin embargo el texto meta traducido automáticamente puede ser interpretado en un caso extremo erróneamente por los receptores en la cultura meta o perder la función comunicativa del texto origen. También la utilización de modismos, colocaciones o lenguaje familiar, como vimos anteriormente, puede dificultar el proceso de traducción automática.

2.2.4.1.2.4 Otros tipos de problemas

Otro tipo de los problemas que se pueden presentar en la traducción automática es la traducción de palabras que no deben ser traducidas. En algunos casos, por ejemplo, en textos técnicos hay cierto tipo de términos como nombres propios, siglas o anglicismos, que están en inglés en el original y no necesitan, ni deben ser traducidos. En una traducción al japonés se refiere a “Apel Computers” como アップルコンピュータ (Lommel et al. 2016) cuando la expresión Inglés debería haberse dejado sin traducir. Aunque este no es propiamente un problema en el texto original, la causa del error se encuentra en el original, por eso lo incluimos en este grupo de problemas.

2. Tecnologías de la traducción, traducción automática y sus problemas

2.2.4.2 Problemas que se presentan dependiendo del motor utilizado

Hasta el momento no hemos hecho una distinción clara entre los problemas que se presentan en la traducción automática dependiendo del sistema o motor que se utilice. Si bien muchos de los problemas que hemos listado hasta ahora se pueden presentar en los dos tipos de sistemas más implementados hasta al momento —motores estadísticos y motores basados en reglas— debido a la naturaleza o programación misma del sistema ciertos problemas tienden a aparecer más cuando se utiliza un sistema y no otro. A continuación clasificaremos los problemas según la frecuencia que se presentan en cada motor.

2.2.4.2.1 Traducción automática basada en estadísticas

Como enunciamos anteriormente los motores estadísticos buscan en el idioma meta la formulación que sea fiel al texto origen y que se ajuste a éste de la mejor manera tanto lingüística, como estilísticamente. Ambos parámetros cualitativos son medidos según probabilidades basadas en un corpus. Es decir, una traducción en la lengua meta es la más adecuada, cuanto más alta sea la probabilidad de que se utilice y es más fiel al original, cuanto más alta sea la posibilidad de que se haya usado como traducción para una oración similar o igual en el texto origen. Sin embargo suelen presentarse problemas que generan errores en la traducción. Como los SMT utilizan para el entrenamiento textos del dominio correspondiente a la traducción se maneja mejor la terminología, pero se presentan problemas gramaticales, como por ejemplo la transferencia correcta de la negación. A continuación listaremos algunos de algunos de los problemas que suelen presentarse en este tipo de motores.

1) Problemas de integridad morfológica. En algunos casos palabras del texto origen no son reconocidas y por lo tanto no se traducen (omisiones). Además puede ocurrir que ciertos verbos no sean identificados como verbos y que sean traducidos como sustantivos. Este tipo de problemas se dan con mayor o menor frecuencia dependiendo de la combinación lingüística.

2) Problemas de ordenamiento sintáctico. Ciertas combinaciones lingüísticas tienen un orden semántico diferente, lo que hace difícil la traducción automática entre los idiomas en cuestión. Es por ello que pueden presentarse problemas en la traducción automática especialmente de oraciones complejas (Schumann y Sivaloganathan 2013, 16).

2. Tecnologías de la traducción, traducción automática y sus problemas

Das Programm **hat**, nachdem es das Update-Prozess abgeschlossen hatte, alle von Benutzer gelieferte Daten, die von der automatische Prüfung als gültig ermittelt wurden, in der Datenbank **geschrieben**.

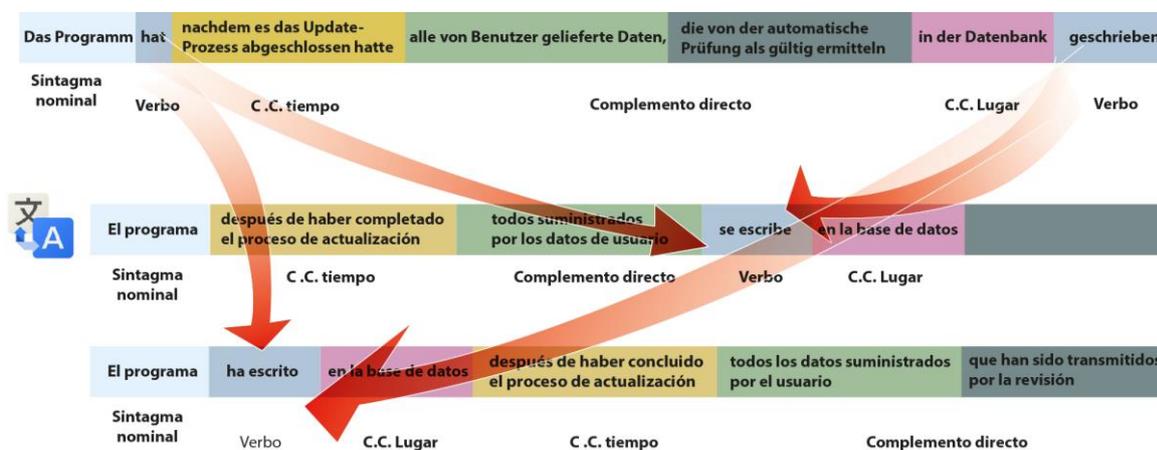


Ilustración 13: Diferencias en la posición del verbo en alemán y castellano

A continuación otro ejemplo menos complejo que el anterior (Aranberri 2015, 33):

Original	<i>Tap "Import" to copy your Android browser favorites</i>
Traducción automática	Toca "Importar" para copiar su navegador de Android favoritos

Tabla 75: Ejemplo de ordenamiento sintáctico incorrecto

En el ejemplo podemos ver que no aunque se intenta se presentan errores en traducción que tienen que ver con el orden sintáctico en inglés.

3) Flexión incorrecta. Puede suceder que se confundan tiempos verbales, que se escriben del mismo modo, es decir, que el infinitivo se traduzca como imperativo, o que un verbo conjugado se traduzca con el infinitivo flexión incorrecta como en el ejemplo siguiente (Aranberri 2015, 33).

Original	<i>Connect your computer to the ZON HUB via Ethernet cable</i>
Traducción automática	Conectar su ordenador a la HUB af a través de cable Ethernet

Tabla 76: Ejemplo de errores en la conjugación de los verbos en los motores SMT

Como enunciamos anteriormente los problemas que se pueden presentar están también determinados por la combinación lingüística que se traduzca. Debido a que el castellano es un

2. Tecnologías de la traducción, traducción automática y sus problemas

idioma con alta inflexión, se presentan muchos errores en la traducción de verbos por ejemplo al inglés, que tienen menos posibilidades. También se presentan errores de tipo léxico al tratar de traducir los verbos ser y estar. El siguiente tipo de errores que más se presenta es la omisión de palabras y el ordenamiento gramatical, en esta combinación es muy frecuente que se presenten problemas al traducir la combinación “adjetivo-sustantivo” ya que en castellano es más común encontrar la combinación “sustantivo-adjetivo”. También se presentan errores en la concordancia de sustantivos, adjetivos y artículos relacionados porque a diferencia del inglés en castellano los artículos y los adjetivos deben ajustarse al género y número del sustantivo. Si cambiamos de dirección, es decir, si se traduce del castellano al inglés el error principal es que se realiza una selección léxica incorrecta, o sea se traducen palabras incorrectamente, porque no se resuelven adecuadamente las ambigüedades. En la tabla a continuación podemos ver una clasificación de errores en estas combinaciones lingüísticas (Vilar et al. 2006):

Type	Sub-type	E-S [%]		S-E [%]	
		FTE	Verbatim	FTE	Verbatim
Missing Words		19.9	26.4	26.0	19.6
	Content Words	7.9	9.9	7.2	4.4
	Filler Word	12.0	16.5	18.8	15.2
Word Order		15.4	11.5	20.4	21.1
	Local Word Order	11.6	4.8	12.7	13.2
	Local Phrase Order	2.1	5.5	6.0	6.9
	Long Range Word Order	1.7	1.1	0.6	1.0
	Long Range Phrase Order	0.0	0.0	1.1	0.0
Incorrect Words		64.4	61.0	50.8	57.3
	Sense	21.9	24.6	28.2	36.8
	Wrong Lexical Choice	13.0	15.4	15.5	21.1
	Disambiguation	8.9	9.2	12.7	15.7
	Incorrect Form	33.9	30.2	9.9	11.7
	Verbs				
	Incorrect Tense	15.1	13.2	7.7	7.8
	Incorrect Person	8.2	8.5	2.2	3.9
	Concordance				
	Incorrect Gender	7.5	4.8	0.0	0.0
	Incorrect Number	3.1	3.7	0.0	3.9
	Extra Words	0.0	2.9	1.1	3.9
	Style	7.9	3.3	9.9	3.9
	Idioms	0.7	0.0	1.7	0.0
Unknown Words		0.3	1.1	2.8	2.0
	Unknown Words	0.3	1.1	1.1	1.5
	Unseen Forms	0.0	0.0	1.6	0.5

Ilustración 14: Comparación de errores presentados en SMT

2. Tecnologías de la traducción, traducción automática y sus problemas

El estudio de Aranberri también es interesante para nuestra tesis porque se ocupa de la traducción de textos técnicos. Al analizar los resultados la autora encuentra que en la combinación lingüística inglés-castellano los errores más comunes se encontraron fueron de tipo léxico (31%), aquí la mayor parte de problemas se encontraron en la traducción de compuestos, que no fueron reconocidos como unidad, así como las traducciones de los nombres de programas o textos que aparecen en el *display*. También se encontraron problemas terminológicos, que en parte se dieron por no haber sido resuelta correctamente algún tipo de ambigüedad. El 29% de los errores encontrados fue de tipo morfo-sintáctico (preposiciones, marcas de subordinación. 18% de los errores tienen que ver con los verbos, principalmente en el modo pero también en el tiempo o la voz. Esto es típico de los textos de instrucción, donde las órdenes dadas con el infinitivo en inglés se pueden traducir en castellano como imperativos o infinitivos. 11% de los errores encontrados tienen que ver con el orden sintáctico, principalmente en la composición de sintagmas nominales de varias palabras. Finalmente 6% de los errores son de puntuación y 5% por palabras no traducidas. La autora da en su artículo una clasificación más detallada y una tabla con algunos ejemplos de los errores encontrados más representativos según las categorías definidas (Aranberri 2015, 33ss).

2.2.4.2.2 Traducción automática basada en reglas

También en la traducción basada en reglas nos encontramos con problemas y errores de traducción. Ya que estos sistemas se basan en el análisis lingüístico del texto original y en una representación abstracta del significado de este, que luego se transfiere al idioma meta, suelen clasificarse problemas de orden terminológico o de contexto.

1) No se traducen todas las palabras. Este tipo de error se presenta en ambos tipos de motores de traducción automática. Causa de ello es en algunas ocasiones que el texto origen puede presentar errores y por ello ciertas palabras (escritas erróneamente) no son traducidas, quedando en el texto meta en el idioma base, o su traducción es equivocada.

2) Problemas con las categorías gramaticales. Estos errores se presentan principalmente con este tipo de motores, aquí se confunden las categorías gramaticales y se incluye la traducción de un verbo en vez de sustantivo o un adjetivo o de un adjetivo en vez de nombre (Eisele 2008).

2. Tecnologías de la traducción, traducción automática y sus problemas

Original	<i>I wish the negotiators continued (adjetivo) success with their work in this important area</i>
Traducción automática	<i>Ich wünsche, dass die Unterhändler Erfolg mit ihrer Arbeit in diesem wichtigen Bereich fortsetzen (verbo)</i>

Tabla 77: Ejemplo de error en la identificación de categorías

3) Flexión de los verbos. También en las traducciones que utilizan este tipo de motores se producen errores en la conjugación de los verbos.

Original	<i>Liegen dem Personal nicht die notwendigen Kenntnisse vor, so ist dieses zu schulen und zu unterweisen.</i>
Traducción automática	Si no tiene los conocimientos necesarios el personal, así se tiene que formar e instruirse esto

Tabla 78: Ejemplo de error en la conjugación de los verbos

Expresiones idiomáticas no se reconocen

Original	<i>We seem sometimes to have lost sight of this fact</i>
Traducción automática	<i>Wir scheinen manchmal Anblick dieser Tatsache verloren zu haben</i>

Tabla 79: Ejemplo de error en la traducción de una expresión idiomática

2.2.4.3 Problemas de la traducción automática en la combinación lingüística alemán-castellano

En la traducción automática en esta combinación lingüística nos encontramos con problemas, que tienen que ver tanto con la traducción automática misma, como con la combinación lingüística. En la revisión de la bibliografía no encontramos estudios que se refieren en particular a los problemas de la traducción automática en nuestra combinación lingüística. Este también fue uno de los motivos que nos animó a iniciar el estudio de la tesis. En el marco metodológico tendremos la oportunidad de ver en un caso concreto, cuáles son los problemas que se presentan en la traducción automática de textos técnicos del alemán al castellano.

2. Tecnologías de la traducción, traducción automática y sus problemas

2.2.4.4 Errores que se presentan en la traducción automática

Si bien algunos de los problemas que nos encontramos en la traducción automática también se encuentran en la traducción humana, los errores que nos encontramos en la traducción automática no son iguales a los que se presentan en la traducción humana. Así como vimos en el capítulo “Errores en la traducción” a partir de la página 40 también se han desarrollado diversas clasificaciones y estudios de los tipos de errores que se presentan en la traducción automática. En los apartados siguientes presentaremos algunas de las que consideramos más importantes para el desarrollo de nuestra tesis.

2.2.4.4.1 Clasificaciones de errores en la década de los ochenta

Existen diferentes aproximaciones al tema de cómo definir si una traducción automática es buena o no. Ya en trabajos realizados a comienzos de los años 80 encontramos algunas tipologías de los diversos problemas que se presentan en la traducción automática. Green (1982) define como “problema” en este contexto la cantidad de esfuerzo físico y mental que se requiere para corregir los errores encontrados. Es decir, para Green la palabra *problema* representa en realidad el esfuerzo cognitivo real en la corrección de la traducción, es decir para el caso de la traducción automática el trabajo de posesición. Según este autor se puede hablar de tres niveles de error en la traducción automática: errores menores, mayores y áreas grises.

Tipo de errores	Ejemplos
Errores menores	Mal uso u omisión del artículo definido.
	Preposición errónea.
	Pronombre personal incorrecto.
	Selección errónea de la traducción, cuando existen varias posibilidades. Por ejemplo cuando para una palabra como <i>dossier</i> (francés) que tiene dos posibles traducciones, carpeta o espaldar, no se escoge la traducción adecuada al contexto.

2. Tecnologías de la traducción, traducción automática y sus problemas

Tipo de errores	Ejemplos
Errores mayores	<ul style="list-style-type: none"> • Traducción palabra-por-palabra de una expresión idiomática. • Aquellos errores que surgen cuando los sistemas identifican incorrectamente partes del texto, como cuando un verbo se traduce como una preposición, por ejemplo: cuando se traduce <i>entre</i> del francés al inglés como <i>between</i> en lugar de <i>enters</i>. • Incapacidad de cambiar verbos activos en el pasivo en inglés.
Áreas grises	Son los errores más difíciles de resolver en el proceso de posesición, ya que no se tiene claridad sobre ellos. En los errores de este tipo se deja a consideración subjetiva del poseedor si deben ser corregidos y hasta qué punto se deben corregir (Green 1982, 182).

Tabla 80: Clasificación de errores de Green

En el mismo año Lavorel (1982) hace también una clasificación de errores apoyándose en un estudio que partía por un lado, de la experiencia del autor como poseedor y por otro, de una encuesta que reflejan las reacciones de seis traductores que realizan la posesición de nueve documentos de carácter principalmente técnico en la combinación lingüística inglés-francés durante un período de dos meses. Los tipos de errores encontrados en la primera parte se dividen en cuatro categorías, que van de los más a los menos problemáticos:

Tipo de errores	Ejemplos		
	Inglés	Traducción automática	Corrección
1. Formas verbales incorrectas	<i>They will be asked to...</i>	<i>Ils seront demandés de</i>	<i>Ils seront invités à</i>
2. Traducción errónea de preposiciones	<i>The delegation agreed to send...</i>	<i>La délégation a consenti d'envoyer</i>	<i>La délégation a consenti à envoyer</i>
3. Traducción literal de expresiones idiomáticas	<i>Mr X gave a progress report on the project</i>	<i>Mr X a donné un rapport d'avancement sur le projet</i>	<i>Mr X rend compte des travaux réalisés dans le cadre du projet</i>

2. Tecnologías de la traducción, traducción automática y sus problemas

Tipo de errores	Ejemplos		
	Inglés	Traducción automática	Corrección
4. Traducción consistente de una palabra a un solo equivalente, cuando el contexto exige que se adacte a otras.	<i>further</i> <i>further work</i> <i>further consideration</i> <i>without further delay</i>	<i>davantage</i>	<i>davantage</i> <i>la poursuite des travaux</i> <i>complément d'examen</i> <i>sans plus tarder</i>

Tabla 81: Clasificación de errores de Lavorel

La segunda parte de la encuesta en la que Lavorel recoge las opiniones de los traductores arrojó otra lista de problemas y opiniones respecto a estos:

1. La falta de una terminología común. La traducción automática no utiliza formulaciones estándar conocidas o no utiliza la expresión correcta en el contexto.
2. La traducción automática no hace ningún esfuerzo por producir propuestas correctas y por eso el poseedor tiene que reformular las propuestas.
3. El estilo de la traducción automática es repetitivo, pero por lo general el poseedor no tiene libertad de cambiarlo.
4. Otros errores mencionados por los entrevistados fueron los términos no traducidos, frases borradas, errores de ortografía y inflexión incorrecta de sustantivos, adjetivos y/o verbos.

En general, se enumeran los mismos tipos de errores en las dos partes del estudio pero en un orden diferente de importancia. Por otra parte los traductores hacen hincapié en la falta de una terminología común, punto que no se enumera en la primera parte. El segundo grupo parece estar molesto porque la traducción automática no puede reproducir el tipo de lenguaje cotidiano de un traductor. La experiencia en el trabajo con traducción automática por un período más prolongado da al autor una mejor idea de lo que se puede esperar y disminuye la apreciación de que no es posible que la traducción automática no sea capaz de producir buenas traducciones. Entre más experiencia se tenga en la posesión, menores son las expectativas y más neutral la apreciación con respecto a la traducción automática (Lavorel 1982, 108).

2. Tecnologías de la traducción, traducción automática y sus problemas

Krings afirma que a pesar de que las clasificaciones de errores de Lavorel y de Green se basan en su propia experiencia práctica y, como tal, tienen una base empírica que impone un cierto respeto, en última instancia, este tipo de clasificaciones son puramente intuitivas y poco sistemáticas, es decir son clasificaciones “ad hoc” (Krings 2001, 56). Las categorías establecidas por estos autores no se definieron con base en criterios operativos, sino se definen exclusivamente con ejemplos. Por lo tanto, no encontramos criterios operacionales que permitan identificar a qué tipo de categorías puede pertenecer otro error que se dé y que no se encuentre entre los citados expresamente por los autores. Krings por su parte también propone una clasificación no centrada en los errores, sino basada en la calidad de una traducción automática. La traducción automática es de calidad baja, si la relación entre el número de errores y palabras es de 1: 2; de calidad media si la relación es de 1:3 y buena si la relación es 1:4. Aunque la diferencia aparentemente no es considerable, este cambio de relación tiene un efecto notable en el proceso de posesición, pues la relación entre el número de errores y la dificultad en el proceso de posesición no es lineal, sino exponencial. Los errores que fueron clasificados como medios, requieren más esfuerzo durante la posesición que los que se clasificaron como malos, ya que estos últimos no tienen que ser corregidos sino desechados.

Loffler-Laurian (1983, 67) realiza también en la década de los ochenta una clasificación más detallada de 12 categorías de errores, partiendo de tres tipos principales: errores en palabras aisladas, errores concernientes a las relaciones y errores que tienen que ver con la estructuración de la información.

Categoría	Errores
Errores en palabras aisladas	1. Vocabulario, terminología.
	2. Nombres propios y abreviaturas.
	3. Relatores, preposiciones: a) En grupos nominales. b) En grupos verbales.
Errores concernientes a las relaciones	4. Determinantes nominales, modificadores verbales.
	5. Tiempos verbales.

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Errores
	6. Voz verbal activa y pasiva; personalización del enunciado.
	7. Expresión de modalidad.
	8. Negaciones.
Errores que afectan la estructura de la información	9. Relaciones lógicas entre los enunciados e introducciones de frases.
	10. Información que se suma o se omite.
	11. Orden de las palabras.
	12. Problemas generales de incidencia.

Tabla 82: Clasificación de errores de Löffler-Laurian

Este tipo de clasificación es más general que las dadas por Green y Lavorel y permite —siguiendo la objeción de Krings que expusimos anteriormente— ser utilizada para el análisis de errores ya que parte de categorías y no de ejemplos.

2.2.4.4.2 Clasificación de LISA

En la década de los 90 encontramos otros tipos de clasificaciones de errores como la de LISA (*Localization Industry Standards Association*). La clasificación fue editada en varias ocasiones y la última versión del modelo (v3.1) puede utilizarse tanto con el software mismo de LISA QA, como emplearse con otras herramientas de localización. La clasificación de LISA se diseñó para ser manejada por editores humanos y la mayor parte de sus categorías no son adecuadas para la evaluación completamente automática pues requieren del juicio del evaluador. La clasificación incluye errores tanto lingüísticos como de otros tipos, por ejemplo errores de maquetación o errores funcionales e incluye una serie de categorías de que tienen que ver con procesos. En realidad no se propago oficialmente como estándar a pesar de que en la industria se utilizó de hecho como tal. Las puntuaciones del modelo se dan generalmente en porcentajes y se calculan con base en un texto de muestra.

2. Tecnologías de la traducción, traducción automática y sus problemas

Documentación	Ayuda	Software
Doc Language (7) <ul style="list-style-type: none"> • Mala traducción • Precisión • Terminología • Lenguaje • Estilo • Región/País • Consistencia Doc Formatting (6) <ul style="list-style-type: none"> • TOC • Índice • Maquetación • Tipografía • Gráficos • Llamadas y leyendas Doc Formatting _Asia (8) <ul style="list-style-type: none"> • TOC • Índice • Maquetación • Tipografía • Gráficos • Llamadas y leyendas • Tamaño doble /sencillo • Marcas de puntuación 	Help Formatting (9) <ul style="list-style-type: none"> • Índice • Maquetación • Tipografía • Gráficos Help Formatting – Asian (4) <ul style="list-style-type: none"> • Índice • (Maquetación) • Gráficos • (Tipografía) • Texto localizable • Hipertexto funcional, saltos, pop-ups 	Software Formating (5) <ul style="list-style-type: none"> • Gráficos • Alineamiento • Apresto • Truncamiento/superposición • Formateo de los caracteres Software functionality testing (5) <ul style="list-style-type: none"> • Texto localizable • Funcionalidad de diálogos • Funcionalidad de menús • Teclas de acceso rápido, aceleradores • Saltos/enlaces
Total 21	Total 8	Total 10

Tabla 83: Categorías de errores de LISA QA v3.1

El modelo presenta además una división en categorías de error de nivel superior en la documentación:

Categorías	Errores
Relacionadas con el idioma en general:	<ul style="list-style-type: none"> • lenguaje • estilo • país
Relacionadas con la documentación:	<ul style="list-style-type: none"> • lenguaje • formato • función
Relacionadas con el software:	<ul style="list-style-type: none"> • lenguaje • formato • función

Tabla 84: Las categorías de errores más importantes en LISA

2. Tecnologías de la traducción, traducción automática y sus problemas

El modelo propone también un sistema de puntuación de los errores según tres niveles de gravedad: los errores menores reciben 1 punto; los errores mayores 5 puntos y los errores críticos 10 puntos. El texto elegido es evaluado manualmente por un revisor: si la puntuación se mantiene por encima de un umbral predefinido, la traducción se considerará aceptable; si cae por debajo del umbral predeterminado, la traducción se considera errónea.

Una desventaja del software de LISA es que solo permite el conteo de los errores, pero no admite el marcarlos para la corrección posterior.

2.2.4.4.3 Clasificación de errores de SAE J2450

SAE J2450 fue originalmente un estándar desarrollado por la *Society of Automotive Engineers (SAE International)* con el fin de satisfacer las necesidades de la industria del automóvil en lo que concierne al control de calidad de las traducciones de todo tipo (humanas, asistidas con sistemas TAO o realizadas empleando sistemas de traducción automática) y en cualquier idioma. No es tan ampliamente utilizado en la traducción, ni en la industria de la localización en general como el modelo de control de calidad de LISA, pero sigue siendo uno de los indicadores de error más comúnmente usados. La última versión fue publicada en agosto del 2005 (SAE International 2005).

Este estándar tiene la ventaja de la simplicidad y se centra estrictamente en la calidad lingüística. Está diseñado para ser utilizado por evaluadores humanos y establece siete categorías principales de errores y dos categorías secundarias; para cada categoría principal que se determinan a su vez dos niveles de gravedad: menor o serio.

Categorías primarias de error	Abreviatura	Categorías secundaria	Peso s m
<i>Wrong Term</i>	WT	Error grave (serious)	5 / 2
<i>Wrong Meaning</i>	WM		5 / 2
<i>Omission</i>	OM		4 / 2
<i>Structural Error</i>	SE		4 / 2
<i>Misspelling Error</i>	SP		3 / 1

2. Tecnologías de la traducción, traducción automática y sus problemas

Categorías primarias de error	Abreviatura	Categorías secundaria	Peso s m
<i>Punctuation Error</i>	PE	Error menor (minor)	2 / 1
<i>Miscellaneous Error</i>	ME		3 / 1

Tabla 85: Clasificación de errores de SAE J2450

Según esta clasificación un error de terminología se presenta si la denominación en el texto meta no corresponde con la terminología especializada o no se emplea de manera consistente. Es decir, se debe usar siempre el mismo término para el mismo objeto o situación, aunque para evitar redundancias y hacer el texto más legible se permite reducir el concepto a su raíz léxica.

Se habla de un error de contenido si el significado del texto meta se diferencia claramente y en gran parte del texto original. Una omisión se presenta cuando no se encuentra en el texto meta una afirmación que se encontraba en el texto original, por ejemplo, porque falta una palabra o parte de una oración. Errores estructurales son aquellos que tienen que ver con la construcción de las palabras, errores gramaticales (conjunciones o adverbios, erróneos, errores en el modo o tiempo de los verbos, errores de coherencia, de caso, preposiciones falsas, uso erróneo de artículos) o errores de sintaxis u orden gramatical. Un error de ortografía se presenta, obviamente, cuando no se cumplen las reglas de la lengua meta; lo mismo que un error de puntuación se da si se infringen las reglas correspondientes o no se utilizan consecuentemente. Los demás errores que no se encuentren en esta categoría se pueden agrupar en el último tipo “otros errores”. Hay que tener en cuenta que SAE-J2450 no tiene en cuenta errores de estilo. O sea que este tipo de errores no se deben incluir en esta última categoría.

Se puede distinguir si un error es serio o menor, cuando el error puede cambiar el sentido del texto meta de tal manera, que puede llevar al usuario —con menor o mayor probabilidad— a realizar una acción errónea que tenga como resultado daños materiales o personales. Mientras que un error menor solo produce una ligera irritación en el usuario, que puede llevar a una acción errónea sin generar daños ni materiales, ni personales. La selección entre los niveles y categorías de errores se determina a través de cuatro “meta-reglas”:

2. Tecnologías de la traducción, traducción automática y sus problemas

- Si se presentan diferentes tipos de errores, se debe elegir el primero de la lista.
- Si el error no se pueden categorizar de manera clara, es mejor categorizarlo como grave. Si se trata de una descripción se puede categorizar como menor, pero si se trata de una instrucción o de informaciones de seguridad es mejor en todo caso categorizarlo como grave.
- No se le debe asignar una puntuación a los errores que se presentan porque había un error en el original, pero se deben marcar como SLT (*Source Language Text*).

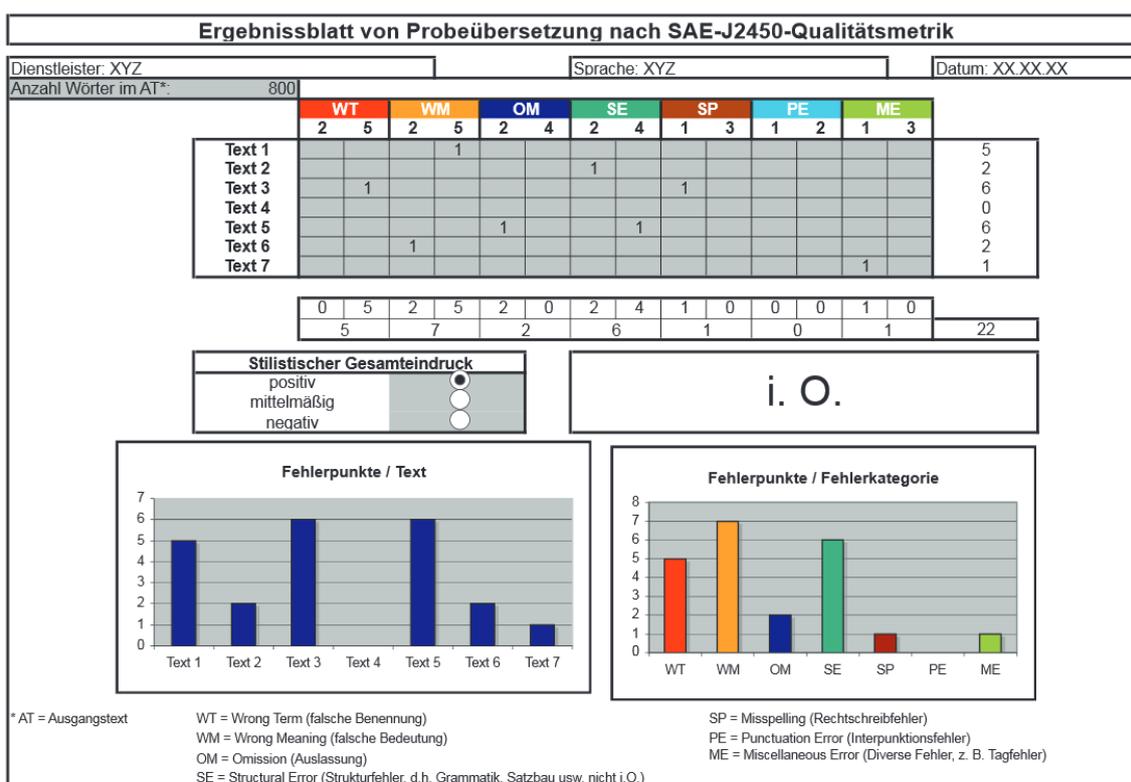


Ilustración 15: Hoja de resultados de un proyecto evaluado con SA-J2450 (Dalla-Zuanna 2010, 23)

Si se presenta un error en el texto meta por un error en el material de referencia tampoco se le debe asignar puntuación, pero se debe marcar como LD (*Legacy Data*).

2.2.4.4 Estudio de TAUS

En un estudio empírico de diferentes modelos de evaluación de calidad comerciales que contaba con el apoyo de TAUS O'Brien (2012) analizó, entre otros puntos, si los modelos del estudio definían categorías de error específicos que pudieran ser útiles para la estimación de calidad y cuáles eran estas categorías.

2. Tecnologías de la traducción, traducción automática y sus problemas

En dicho estudio comparativo realizado en el 2011 una pequeña muestra de miembros de TAUS fueron invitados a colaborar en el proyecto con el objeto de realizar una evaluación de once modelos de estimación de calidad de traducción automática. Las empresas fueron seleccionadas por TAUS teniendo como criterios que representaran diferentes sectores de actividad y que estuvieran interesados en nuevos métodos de evaluación de la calidad. Por motivos de confidencialidad no se dieron a conocer los nombres de los participantes pero se mencionó que ocho de los modelos incluidos estaban en uso activo. Seis empresas operan en el ámbito TI multinacional, una trabaja en el campo de la electrónica de consumo y otra en el desarrollo de dispositivos médicos. Para participar en el estudio empresas también tenían que estar dispuestas a compartir sus modelos de evaluación de calidad.

Los once modelos se pudieron clasificar en dos categorías generales. La primera categoría era evaluar la calidad de la traducción, desde el punto de vista de los errores que se encuentren en el texto traducido. Diez de los modelos de estimación de calidad encajaban en esta categoría. En la segunda categoría se consideraba la calidad de la traducción de una manera más amplia, teniendo en cuenta la prestación de servicios y las competencias, instrumentos y procedimientos necesarios para producir una traducción de alta calidad. Este es un modelo de estimación de calidad orientado al proceso. Este último modelo no enumera específicamente errores de traducción y por ello no se puede comparar directamente con los demás. Sin embargo, la lista los problemas que los traductores deben tener en cuenta al traducir (por ejemplo la terminología) y estos temas en términos generales pueden ser asimilados a algunas de las categorías de errores generales identificados en los otros diez modelos de estimación de calidad.

En el estudio se encontró que la categoría de errores más importante para la mayoría de los modelos era *Language*. En los errores en el lenguaje, se encontraron las siguientes subcategorías:

- *Language* (relevante para 10 de los 11 modelos).
- *Terminology* (relevante para 10 de los 11 modelos).
- *Accuracy* (relevante para 9 de los 11 modelos).
- *Style* (relevante para 7 de los 11 modelos).

2. Tecnologías de la traducción, traducción automática y sus problemas

Las categorías menos relevantes fueron:

- Las normas regionales o *country standards* (relevante para 4 de 10 modelos).
- *Mistranslation* (relevante para 3 de 10 modelos).
- *Consistency* (relevante para 3 de 10 modelos).

Como se puede ver en la lista anterior ciertas categorías de errores son comunes en muchos de los modelos de estimación de calidad que se incluyeron en el estudio comparativo de evaluación, pero se requiere de un análisis más detallado de lo que implica cada categoría para determinar si las categorías específicas de los tipos de errores son comparables en términos generales o no. La autora analizó las cuatro principales (sub)categorías de errores definidos como errores de *Language*. Este estudio reveló que algunos modelos utilizaron términos diferentes para denominar la misma categoría, pero al examinar las definiciones, se pudo definir si se estaba hablando de la misma categoría o si se trataba de una diferente. Solo algunos modelos definieron errores en la autoedición (DTP) y en la interfaz de usuario. Finalmente O'Brien encontró las siguientes categorías de errores:

1) Errores del lenguaje

No parece haber un consenso general entre los participantes del estudio en cuanto a lo que se entiende por *Language* como categoría. Nueve de los modelos analizados incluían la gramática en sus definiciones y siete la sintaxis. Otros criterios dominantes fueron la puntuación (seis referencias) y la ortografía (siete referencias); sin embargo hubo tres modelos que definieron puntuación y ortografía como criterios independientes del 'Lenguaje'. También se encontraron algunos criterios menos comunes como errores tipográficos, fluidez y referencias culturales.

a) Errores en la terminología. Diez de los once modelos se refieren específicamente a la terminología como una categoría de error. Hubo un consenso general en todos los modelos de evaluación de calidad que incluían la categoría "terminología" y es que los errores en esta categoría implican 1) la falta de coherencia con un glosario específico del cliente (u otros materiales de referencia); 2) la falta de coherencia en el uso del término específico de la terminología de la industria y 3) la falta de coherencia en el uso del término mismo. Además, tres de

2. Tecnologías de la traducción, traducción automática y sus problemas

los modelos de evaluación de calidad mencionan el uso inadecuado en el contexto como un criterio para determinar un error de terminología.

b) Precisión. Nueve de once modelos de evaluación de calidad se refirieron a este concepto. Siete incluyeron omisiones, adiciones innecesarias y referencias cruzadas inexactas en sus definiciones de precisión y también incluyeron errores de significado.

c) Estilo. Siete de once modelos de evaluación de calidad mencionan 'Estilo' como una categoría de error, un modelo no tuvo en cuenta el estilo y otro solo menciona el estilo como algo que los traductores (y revisores) tienen que tener en cuenta. Cuatro de siete modelos listan la "falta de consistencia con la guía de estilo del cliente" como criterio para clasificar un error de estilo. Aparte de esto hay poco consenso sobre lo que constituye un error estilístico. Algunos criterios incluyen el tono, registro, variantes lingüísticas, argot, traducciones literales y la sintaxis torpe. De los cuatro primeros errores que hacen parte de la categoría 'Lenguaje', 'Estilo' es el que se da menos consenso entre los modelos.

2) Errores de Desktop-Publishing o autoedición

A pesar que modelos de estimación de calidad públicos como el J2450, que describimos en la página 133, descartan cualquier tipo de error que no esté relacionado con el lenguaje y únicamente EN15038 menciona la autoedición, pero no enumera los tipos de error que se dan en el proceso, O'Brien lista las categorías de errores en autoedición que se mencionan los otros modelos de estimación de calidad:

- Maquetación (relevante para 5 de 11 modelos).
- Formato (relevante para 4 de 11 modelos).
- Gráficos (relevante para 3 de 11 modelos).
- Índice (relevante para 2 de 11 modelos).
- Tabla de contenido (relevante para 2 de 11 modelos).
- Funcionalidad, p.ej. etiquetas o variables (relevante para 2 de 11 modelos).

3) Errores en la interfaz de usuario

El modelo LISA QA (v.3.1), véase página 131, contiene una lista detallada de errores de autoedición y errores de traducción relacionados con la interfaz de usuario. O'Brien enu-

2. Tecnologías de la traducción, traducción automática y sus problemas

mera las categorías de errores en la interfaz de usuario que se mencionan más frecuentemente. La autora recalca que esta lista no es exhaustiva, ya que es poco probable que en ella se listen todos los errores que se pueden presentar relacionados con la traducción de software:

- Integridad de texto localizado (relevante para 3 de 11 modelos).
- *Alignment* (relevante para 2 de 11 modelos).
- *Layout* (relevante para 2 de 11 modelos).
- *Truncation* (relevante para 3 de 11 modelos).

En la tabla a continuación podemos ver las categorías y los subcategorías de errores correspondientes, con una breve descripción.

Categoría	Errores
Errores del lenguaje	Terminología: <ul style="list-style-type: none"> • Falta de consistencia con un glosario específico. • Falta de consistencia en el uso terminológico general. • Falta de consistencia en la utilización interna del término.
	Precisión: Omisiones, adiciones innecesarias y referencias cruzadas inexactas, errores de significado.
	Estilo: <ul style="list-style-type: none"> • Falta de consistencia con la guía de estilo del cliente • Tono, registro, variantes lingüísticas, argot, traducciones literales y la sintaxis torpe.
Errores de autoedición	<ul style="list-style-type: none"> • Maquetación. • Formato. • Gráficos. • Índice. • Tabla de contenido. • Funcionalidad (etiquetas, variables).
Errores en la Interfaz	<ul style="list-style-type: none"> • Integridad de texto localizado. • Alienación. • Maquetación o Layout. • Truncation.

Tabla 86: Clasificación de errores del estudio realizado por O'Brien

2. Tecnologías de la traducción, traducción automática y sus problemas

2.2.5 *Métodos de evaluación de la calidad de la traducción automática*

Otro proceso importante para el desarrollo de la traducción automática es el control mismo de la calidad de la traducción. Los métodos de control de calidad son importantes también para comparar los diferentes motores de traducción automática, para medir su progreso a través del tiempo y de esta manera contribuir tanto al desarrollo de los sistemas mismos como a la evaluación de las traducciones. Es difícil determinar con qué criterios se puede definir si una traducción automática es o no de buena calidad. El mismo concepto de calidad es relativo y depende de la expectativa que se tenga y del encargo de traducción. Incluso la gravedad de los errores no se puede generalizar, pues depende del tipo de persona que reciba o evalúe el texto traducido automáticamente: un error no es igual de grave si el texto va a ser poseído o si el texto solo se ha resultado de un *gisting* que está recibiendo un lector que no tiene conocimientos sobre el idioma meta y no puede identificarlo como una mala traducción (Specia 2013).

Es por ello que pregunta relevante es cómo medir la calidad de una traducción automática. Existen métricas manuales que se dedican 1) al conteo de errores, la clasificación de estos y de su aceptabilidad, 2) evaluaciones basadas en la fluidez y/o adecuación de la traducción automática con respecto al original y 3) otras métricas que se centran en la tarea misma, como las evaluaciones de productividad, de satisfacción del usuario, de comprensión de lectura. Así como para el conteo de errores nos encontramos con métricas automáticas que se apoyan o tienen referencia en traducciones humanas (BLEU, METEOR, TER, etc.), también nos encontramos con otras que no utilizan ninguna traducción de referencia (*Quality Estimation*). A diferencia de las métricas estándar de evaluación de la traducción automática, las métricas de *Quality Estimation* no tienen acceso a traducciones (humanas) de referencia; su aplicación está dirigida a los sistemas de traducción automática en uso y se utilizan para decidir qué segmentos deben ser revisados por un traductor y qué tan fiable es la traducción que obtiene el lector. También ayudan a estimar el esfuerzo que se necesita para la posesión; facilitan el proceso de selección entre traducciones alternativas, si se han utilizado diferentes motores y ayudan a decidir si la traducción se puede utilizar para la autoformación o entrenamiento de los sistemas mismos de traducción automática (Specia et al. 2013, 79).

2. Tecnologías de la traducción, traducción automática y sus problemas

El desarrollo de *Quality Estimation* se inició en la década de 2000 inspirado en las puntuaciones de confianza utilizadas en los métodos de reconocimiento de voz y se denominó por ello mismo *confidence estimation*. En el 2004 se realizó en la Universidad John Hopkins un taller sobre el tema (Blatz et al. 2004) que también tuvo como objetivo el evaluar métricas automáticas de estimación de calidad como BLEU (*Bilingual Evaluation Understudy*) (Papineni et al. 2002) y WER (*Word Error Rate*).

Las métricas manuales que miden adecuación y fluidez utilizan escalas de puntos. Por ejemplo para *Accuracy* se otorga la mejor puntuación si todo el significado del fragmento original se encuentra en el segmento traducido; un valor menor si la mayor parte se expresa y la menor puntuación si no se trasmite el sentido original. Para medir la fluidez también se utilizan escalas de puntos que funcionan de la misma manera, la mejor puntuación se le da a una traducción que sea tan fluida como si fuera un texto escrito por un hablante del idioma; si se aproxima a la fluidez, si no es muy fluida o si es poco fluida se le asigna una puntuación menor y finalmente, si no es fluida y no tiene sentido tendrá la menor puntuación posible.

2.2.5.1 *Métodos y métricas de evaluación sin intervención humana*

La idea básica de estas métricas es comparar la salida del sistema de traducción automática con una buena “referencia” de traducción (normalmente humana) y determinar qué tan cerca se encuentra la salida de la traducción automática a la traducción de referencia. Las ventajas de estos métodos son que 1) no se necesitan hablantes bilingües y 2) que pueden ser utilizados durante el desarrollo de un sistema para probar los cambios. Pero también tienen desventajas ya que las métricas actuales no son todavía lo suficientemente detalladas y no distinguen bien las diferencias en los sistemas. Además la puntuación dada a los segmentos individuales a menudo no es muy fiable. Para entender mejor estas afirmaciones veamos a continuación algunas de las métricas automáticas más utilizadas.

BLEU

Bilingual Evaluation Understudy (Papineni et al. 2002) fue una de las primeras métricas utilizada y fue implementada por IBM. BLEU un estándar que parte de la premisa de que una traducción tiene mayor calidad en tanto más se asemeje a otra, que tendrá el rol de referencia y que se supone que es correcta. BLEU puede calcularse utilizando más de una traduc-

2. Tecnologías de la traducción, traducción automática y sus problemas

ción de referencia. En otras palabras, aquí no se comparan la traducción directamente con el original, sino con la referencia. Lo que implica en algunos casos que un valor BLEU alto no significa que la traducción sea buena. Entre más referencias de traducciones humanas se tenga, la puntuación será más precisa. Una desventaja de BLEU es que no le da valor a las propuestas cercanas, por ejemplo cuando se emplean sinónimos.

$$\text{BLEU} = \text{BP}(\text{tran}, \text{ref}) \times \prod_{i=1}^4 \text{precision}_i$$

$$\text{BP}(\text{tran}, \text{ref}) = \min \left(1, \frac{\text{length}(\text{trans})}{\text{length}(\text{ref})} \right)$$

$$\text{precision}_i = \frac{\# \text{common segments of length } i}{\# \text{possible segments of length } i}$$

Ilustración 16: Algoritmo de definición de BLEU (Sánchez-Martínez 2014)

Utilizaremos el siguiente ejemplo de traducción automática y una traducción de referencia para comparar los resultados de algunas de las métricas:

TA	Israeli officials responsibility of airport safety
Referencia	Israeli officials are responsible for the airport security

Tabla 87: Ejemplo para el cálculo de BLEU (Sánchez-Martínez 2014)

BLEU tendría en este caso el siguiente valor:

$$\text{BLEU} = \frac{6}{7} \times \left(\frac{3}{6} \times \frac{1}{5} \times \frac{0}{4} \times \frac{0}{3} \right) = 0$$

Mientras que otra salida similar tendría un valor muy diferente:

TA	airport security Israel officials are responsible
Referencia	Israeli officials are responsible for the airport security

Tabla 108: Cambios en el cálculo de BLEU

2. Tecnologías de la traducción, traducción automática y sus problemas

Aquí el valor sería:

$$\text{BLEU} = \frac{6}{7} \times \left(\frac{6}{6} \times \frac{4}{5} \times \frac{2}{4} \times \frac{1}{3} \right) = 0.1143$$

Koehn (2010, 229) le afirma que esta métrica tiene varios problemas, por ejemplo el que ignora la importancia relativa de las diferentes palabras ya que en un oración algunas palabras son importantes más que otras, por ejemplo una negación. Además con BLEU no se tiene en cuenta la coherencia gramatical general. Además las puntuaciones de BLEU no son fáciles de interpretar, porque dependen de muchos factores, tales como el número de traducciones de referencia, el par de idiomas implicado, el dominio, etc.

NIST Open Machine Translation (OpenMT) Evaluation

El NIST creó una variación de BLEU que le da un mayor peso a los ngramas poco frecuentes. El objetivo del *NIST Open Machine Translation Evaluation* (NIST 2015) es el promover la investigación y ayudar a avanzar en el área de la traducción automática, para que el futuro la salida de la traducción automática sea una traducción adecuada y tan fluida como el original.

La evaluación de la traducción automática se inició en 2001 como parte del programa de DARPA MAREAS. En su forma actual, las evaluaciones son impulsadas y coordinadas por el mismo NIST (*National Institute of Standards and Technology*) como “NIST OpenMT”.

METEOR

Con *METEOR Metric for Evaluation of Translation with Explicit Ordering* (Denkowski y Lavie 2014) se evalúan los resultados de las traducciones automáticas alineándolas a una o más traducciones de referencia. Es por lo tanto una medida de evaluación que cuenta el número de coincidencias exactas de palabras entre la salida del sistema de traducción automática y la referencia. La puntuación se da de acuerdo al grado de similitud que tenga la frase con la referencia. Los alineamientos se hacen con base en traducciones exactas, sinónimos y/o paráfrasis. La métrica incluye varios parámetros libres que se sintonizan para emular varias tareas de juicio humano, incluyendo la adecuación a NIST. METEOR ha extendido el soporte (coincidente paráfrasis y parámetros ajustados) para los siguientes idiomas: inglés, checo,

2. Tecnologías de la traducción, traducción automática y sus problemas

alemán, francés, español y árabe. METEOR está implementado en Java puro y no requiere instalación o dependencias para anotar salida de la traducción automática. En promedio, las hipótesis se califican a un ritmo de 500 segmentos por segundo. METEOR demuestra una alta correlación con los juicios humanos en evaluaciones independientes como EMNLP WMT 2011 y las métrica NIST MATR 2010.

El principal inconveniente de METEOR es que su método y fórmula para calcular la puntuación es mucho más complicada que la de BLEU. Se tienen en cuenta muchos más parámetros (Koehn 2009, 228)

TER

TER se calcula como el número de ediciones que un humano tendría que llevar a cabo para cambiar una salida de un sistema de traducción automática, de tal manera que esta coincida con la traducción de referencia. Ediciones son inserciones y omisiones, así como "cambios" en la secuencias de palabras continuas. TER puede tener en cuenta varias traducciones de referencia y la distancia de edición se calcula con respecto a uno de estas traducciones (Babych 2014).

$$\text{TER} = \frac{\text{\#edits}}{\text{average \# of reference}}$$

HTER

Human-mediated Translation Edit Rate (Snover et al. 2006) es una medida semiautomática más compleja en la que los humanos no califican en primer lugar las traducciones directamente, sino realizan una nueva traducción de referencia que está más cerca de la salida de traducción automática pero conserva la fluidez y el significado de la referencia original. El algoritmo oficial realiza una comparación entre la salida de traducción automática y la versión corregida de esta salida. Los resultados de la comparación cuentan el número de inserciones, deleciones, sustituciones y cambios (por ejemplo reordenación). La puntuación HTER final es el número total de ediciones dividido por el número de palabras de referencia (Avramidis 2014, 302).

2. Tecnologías de la traducción, traducción automática y sus problemas

$$\text{HTER} = \frac{\#insertions + \#dels + \#subs + \#shifts}{\#reference\ words}$$

WER

La tasa de error de palabra emplea la distancia de edición de Levenshtein, es decir calcula el número mínimo de inserciones, omisiones y sustituciones necesarias para que coincidan dos secuencias (Koehn 2009, 224-25). Sin embargo, una distancia de edición estándar (que fue desarrollada para el área de reconocimiento de voz automático) se considera que es demasiado simplista para la traducción automática. La razón es que la traducción legítima varía a menudo dependiendo de las diferencias en el orden de las palabras y frases sin grandes cambios en el significado, y WER penaliza tales reordenamientos, al mismo nivel que el uso de palabras equivocadas en un lugar y la inserción de palabras redundantes o espurias en otro lugar.

$$\text{WER} = \frac{\#subs + \#dels + \#insertions}{\#reference\ words}$$

El valor WER se calcularía —teniendo en cuenta la traducción y la referencia que utilizamos anteriormente para calcular BLEU— de la siguiente manera:

Traducción automática	the Israeli officials responsibility of airport safety
Referencia	Israeli officials are responsible for the airport security

Tabla 88: Ejemplo para el cálculo de WER (Sánchez-Martínez 2014)

$$\text{WER} = \frac{3 + 3 + 0}{8} = 0,75$$

En esta métrica las palabras se cuentan si están en el mismo orden.

PER

PER o *Position-independent error rate* mide el porcentaje de palabras en la traducción de referencia que no aparecen en la traducción automática sin posesición, sin tener en cuenta el orden de las palabras.

2. Tecnologías de la traducción, traducción automática y sus problemas

$$\text{PER} = 1 - \frac{\#ok - \max(0, \text{length}(\text{trans}) - \text{length}(\text{ref}))}{\text{length}(\text{ref})}$$

Si tenemos en cuenta el mismo ejemplo anterior, el valor PER se calcularía de la siguiente manera:

Traducción automática	the Israeli officials responsibility of airport safety
Referencia	Israeli officials are responsible for the airport security

Tabla 89: Ejemplo para el cálculo de WER (Sánchez-Martínez 2014)

$$\text{PER} = 1 - \frac{4 - \max(0, 7 - 8)}{8} = \frac{1 - 4}{8} = 0.5$$

2.2.5.2 Métodos de evaluación con intervención humana

Debido a que las métricas automáticas son muy abstractas y además pueden arrojar resultados diferentes depende de las referencias que se utilicen, un mismo segmento original puede tener varias traducciones correctas. Otra razón para este debate es que los indicadores automáticos se utilizan casi exclusivamente por investigadores de traducción automática estadística, y sus afirmaciones son a menudo cuestionadas por los profesionales de la traducción automática. La otra opción es la evaluación manual realizada por uno varios evaluadores humanos. A continuación presentaremos algunos de estos intentos.

Adequacy/Fluency Guidelines

Evaluaciones que tienen en cuenta los criterios de adecuación y/o fluidez se emplean regularmente para calificar la calidad de la traducción automática y también son útiles para la evaluación de la traducción humana en determinados contextos. Estos métodos son según TAUS(2014) menos costoso y son más rápidos de implementar que un enfoque de tipología de errores y pueden ayudar a centrarse en la evaluación de los atributos de calidad que son más relevantes para ciertos tipos de contenido y propósitos específicos.

La idea central es proporcionar directrices que permitan adoptar enfoques estándar, asegurando un lenguaje compartido entre aquellos que hacen el encargo de traducción, quienes lo realizan y quienes lo evalúan y por supuesto, el reducir los costes del proceso.

2. Tecnologías de la traducción, traducción automática y sus problemas

Como adecuación se define qué tan bien se expresa el significado en la traducción, es decir, si se transmite el sentido en la traducción. Para medir la adecuación se propone una escala de cuatro puntos, con la que se puede clasificar que tanto se transmite en la traducción el significado del original: todo, la mayoría, un poco o nada.

Como fluidez se mide hasta qué punto una traducción es correcta (gramaticalmente), no tiene faltas de ortografía y refleja el uso de lenguaje natural o intuitivo de un hablante nativo. Aquí también se propone una escala de cuatro puntos: si la fluidez es perfecta, buena, si la traducción no es fluida o si es incomprendible

SSEER (Subjective Sentence Error Rate)

En este tipo de sistema más simple el evaluador debe calificar la traducción de una frase dándole la puntuación entre 0 y 10. Una calificación de 0 se le debe asignar a una traducción perfecta, mientras que una de 10 es la puntuación que debe recibir una traducción sintáctica y semánticamente incorrecta. Posteriormente se convierte la puntuación a un porcentaje.

El mayor problema lo demuestra esta técnica es su naturaleza subjetiva, ya que dos personas que evalúen el mismo texto podría obtener resultados bastante diferentes. Para resolver este problema se propone realizar varias evaluaciones. Otro inconveniente es que no se tiene en cuenta las diferentes longitudes de las frases. La puntuación de una frase que tenga 100 palabras tiene el mismo impacto en la puntuación total que una frase o segmento con una sola palabra (Nissen et al. 2000).

IER (Information Item Error Rate)

Para poder evaluar largas cadenas o segmentos compuestos de partes correctas y partes erróneas se creó este sistema de evaluación. Para ello se introdujo el concepto de Ítems o elementos de información. Las frases se dividen en segmentos de palabras. Cada elemento de la frase se marca con “OK”, “error”, “sintaxis”, “significado” u “otros”. IER se calcula como el porcentaje de los ítems mal traducidos —es decir, aquellos que no se marcaron como “OK” (Nissen et al. 2000).

2. Tecnologías de la traducción, traducción automática y sus problemas

Métricas basadas en tareas

Estas métricas se basan, por ejemplo, en el análisis de productividad, es decir, en cuánto se demora el proceso de posesición. Para ello se mide la distancia de teclado, las pulsaciones, para determinar si es más recomendable realizar la posesición que una traducción humana de cero, comparando el tiempo de traducción de cero con el tiempo de la posesición en segundos. Para resolver este dilema se hacen preguntas incluso a los usuarios finales de la traducción automática, tales como si pudo resolver un problema leyendo la traducción automática, o si está satisfecho con la información recibida. Aquí se mide la calidad de la traducción si el texto es comprendido (adecuado) y por el tiempo que llevo realizar el test.

2.2.5.3 Perspectivas actuales de evaluación

A pesar de que históricamente el área de evaluación de la traducción automática comenzó como un campo derivado en el desarrollo de la traducción automática y estaba dirigido principalmente a la medición de las mejoras de los sistemas y sus características durante el ciclo de desarrollo, ahora se ha convertido en un campo independiente con un conjunto más amplio de objetivos y partes interesadas, que incluye no solo la comunidad de desarrolladores de sistemas de traducción automática, sino también traductores, gerentes de proyectos de localización, inversores, etc. Incluso se tiene en cuenta la opinión de usuarios, que tienen que tomar decisiones sobre los sistemas de traducción automática y evaluar si dichos sistemas pueden satisfacer sus necesidades específicas, por ejemplo, desde el punto de vista de la utilidad, la calidad del texto, el esfuerzo de posesición, el ahorro de tiempo y coste en el flujo de trabajo de traducción colaborativa, o para el uso de la producción de la traducción automática inédita solo para la comprensión de textos, etc. (Babych 2014, 465).

TAUS realizó en el 2011 una evaluación comparativa de diferentes modelos de evaluación existentes (O'Brien et al., 2011) que arrojó como resultado que las métricas existentes hasta esa fecha eran relativamente rígidas: las categorías de errores y de sanciones aplicadas, los límites entre acierto/error etc., eran los mismos sin importar que parámetros comunicativos estuviesen implicados. Otro punto crítico lo constituyó el hecho de que los modelos también eran tan detallados que su aplicación implicaba mucho tiempo y solo permitía la evaluación de una pequeña muestra de palabras.

2. Tecnologías de la traducción, traducción automática y sus problemas

En los últimos años ha crecido el interés por un cambio en los tipos estáticos y normativos de evaluación, ya que los métodos actuales, con o sin intervención humana, requieren mucho tiempo y no tienen en cuenta aspectos de importancia como el tipo de contenido, las necesidades de los usuarios, la función comunicativa, el contexto o la manera en que se realizó la traducción (O'Brien 2012).

2.2.5.3.1 *Dynamic Quality Framework (DQF)*

Es por ello que TAUS desarrolló el *Dynamic Quality Framework (DQF)* con el fin de mejorar la evaluación humana del contenido traducido. La plataforma DQF consiste en una rica base de datos sobre evaluación de calidad con las mejores prácticas, informes, plantillas y una serie de herramientas para evaluar las traducciones hechas tanto por traductores humanos, como por motores de traducción automática. Las herramientas permiten a los evaluadores comparar traducciones, evaluar su precisión y fluidez, medir la productividad de la posesión y marcar los segmentos traducidos con base en una tipología de errores. El asistente de perfiles de contenido permite a los usuarios seleccionar los métodos de evaluación de mejor ajuste a sus necesidades (Görög 2014, 444).

El proyecto se inició en enero del 2011 con la participación de más de cincuenta empresas y organizaciones que incluían clientes, proveedores de servicios y de tecnologías de traducción e instituciones académicas; desde el 2014 DQF hace parte de la plataforma de evaluación de TAUS.

DQF es un enfoque evaluativo más flexible que se basa en tres parámetros: *Utility*, *Time* y *Sentiment (UTS)*. Como, *Utility* se define la importancia relativa de la funcionalidad del contenido traducido; bajo *Time* se estima el lapso requerido para la traducción y como *Sentiment* se pretende determinar la importancia del impacto de una mala traducción en la imagen de una marca, es decir, qué tan perjudicial podría ser para un cliente que el contenido estuviera mal traducido.

Además este modelo dinámico de estimación de calidad tiene en cuenta el canal de comunicación para la evaluación y define tres canales:

2. Tecnologías de la traducción, traducción automática y sus problemas

- B2C: Business-to-Consumer
- B2B: Business-to-Business
- C2C: Consumer-to-Consumer.

Estos canales son importantes porque determinan el grado de evaluación a que se deben someter los textos. Por ejemplo, para C2C, categoría en la que se incluyen los contenidos generados por usuarios en varios idiomas, que son recibidos por otros usuarios en (otros) varios idiomas (por ejemplo, tweets, blogs, mensajes en los foros, etc.), no se requiere un modelo de evaluación tan estricto, como para textos que se encuentran en las categorías B2B o C2C. Es por ello que se afirma que modelo de calidad no puede estar separado del canal de comunicación (O'Brien, 2012).

El modelo es dinámico porque para cada tipo de texto o proyecto el impacto de los parámetros *UTS* puede ser mayor o menor y por ello se permite su ajuste. Por ejemplo, *Utility* es un parámetro que tiene más importancia en la traducción de una interfaz de usuario o del contenido del sitio web, que si se trata de una traducción de material de audio/vídeo. *Time* es un criterio que tiene mayor importancia en la traducción de material de marketing y de los contenidos de medios sociales que la documentación de usuario, material de capacitación, ayuda en línea y contenido de audio/vídeo para los que la importancia de este factor tendría un valor medio. *Sentiment* es más relevante en la traducción de material de marketing y del contenido del sitio web, que en la traducción del material de formación o de la ayuda en línea, en los que un error en la traducción no tendría un impacto tan negativo.

Teniendo en cuenta estas apreciaciones DQF propone que las calificaciones UTS se deben asignar a modelos de mejora de la calidad específicos para así poder recomendar el modelo más adecuado de acuerdo con las necesidades de cada usuario. Esta es la base de la herramienta de perfilado contenida en TAUS. Una vez que el usuario eha perfilado el contenido traducido y se haya seleccionado el tipo de evaluación más adecuada, el proyecto de evaluación se puede configurar con las herramientas de DQF. Las herramientas DQF se desarrollaron pensando en un usuario que no tenga conocimientos técnicos.

2. Tecnologías de la traducción, traducción automática y sus problemas

2.2.5.3.2 *Multidimensional Quality Metrics (MQM)*

MQM proporciona un marco para describir y definir indicadores de calidad que se pueden utilizar para evaluar textos traducidos y para identificar problemas específicos en dichos textos. También proporciona un marco sistemático para describir las métricas de calidad basadas en la identificación de las características textuales que permiten evaluar la calidad de las traducciones. Este modelo tiene sus raíces en el modelo LISA QA pero fue diseñado desde el principio para ser flexible y trabajar con otras normas con el fin de integrar la evaluación de la calidad en todo el ciclo de producción de documentos (Lommel, Uszkoreit, y Burchardt 2014).

El componente central de MQM es una lista jerárquica de tipos de problemas, que se estableció después de un examen cuidadoso de las métricas de evaluación de calidad existentes y de los problemas encontrados por las herramientas de control de calidad automáticas. Los temas fueron restringidos a solo los relacionados con el lenguaje y el formato, dejando de lado las cuestiones del modelo LISA QA que tienen que ver con la calidad del proyecto (por ejemplo, asuntos relacionados con la entrega del proyecto).

La jerarquía de errores que explicaremos más adelante (véase página 154) es una parte esencial de MQM. Cada nodo (incluidos los alto nivel como precisión y fluidez) pueden definirse como un tipo de problema y los derivados de un problema representan casos específicos de la cuestión principal. Esto permite que con una métrica MQM se puedan declarar a varios niveles de detalle. Para la evaluación de la calidad de uso general, por ejemplo, una métrica relativamente general es probable que necesite mucho tiempo para el diagnóstico de problemas en un sistema de traducción automática que una métrica mucho más detallada que se centra en cuestiones específicas relacionadas con el desarrollo del sistema, sería más apropiada.

MQM aborda el problema que se ve en las métricas de una talla única para todo (como el control de calidad del modelo de LISA) mediante la definición de un modelo que declara múltiples métricas en lugar de una sola métrica. Proporciona un vocabulario estandarizado para declarar problemas, sin embargo, permite métricas para ser comparadas y, donde hay solapamiento, los resultados también pueden compararse.

2. Tecnologías de la traducción, traducción automática y sus problemas

Aunque MQM apoya un enfoque detallado en la evaluación de la calidad, las mejores prácticas para la evaluación humana se realizarían con una métrica cuya complejidad sea ajustada a los requerimientos de la tarea de evaluación especial, ya que por ejemplo es más difícil para los correctores distinguir entre categorías muy detalladas que entre categorías de nivel superior. Es decir, si la evaluación solo necesita saber si hay errores gramaticales en un texto traducido, no sería necesario incluir en la evaluación todos los subtipos de la gramática; Sin embargo si se está utilizando MQM para evaluar la salida de un sistema de traducción automática si es importante entender los tipos precisos de los problemas encontrados en el sistema de traducción automática y así permitir procesos posteriores para decidir si debe o no convertir los resultados a tipos más generales).

MQM introduce además el concepto de *Verity*. Con este concepto se pretende describir y evaluar el fenómeno de referencia extralingüística de los textos, que se ve particularmente en ciertos textos, por ejemplo, en textos legales escritos para un tipo de legislación en particular. No existe un consenso general acerca de si este aspecto se deba o no considerar como parte de la traducción ya que es una noción extra-lingüística. Además *Verity* se ocupa de cuestiones que están fuera de las nociones tradicionales de precisión y fluidez; y por ello en muchos casos no se puede evaluar simplemente con la lectura de la traducción o comparándolo con la fuente. Para evaluar la veracidad de una traducción se tiene que examinar el texto en su entorno, en el contexto cultural o en el medio correspondiente, ya que una traducción a pesar de ser fluida y precisa podría ser inadecuada. No solo textos jurídicos, también otro tipo de textos pueden necesitar ser modificados después del proceso de traducción; por ejemplo la descripción de las opciones disponibles de ciertos productos en ciertos mercados o ciertas ofertas promocionales.

El foco de atención en MQM se centra en la parte analítica y definatoria, es decir en el identificar los errores específicos en una traducción para poder cuantificarlos. En este tipo de métodos analíticos los errores son vistos como casos desviaciones en un texto ya que no cumple plenamente con las especificaciones y los “errores” que no conducen a una desviación de las especificaciones no son contados como errores. Por ejemplo, si las especificaciones para un manual de servicio definen que los problemas estilísticos no son importantes, un estilo

2. Tecnologías de la traducción, traducción automática y sus problemas

pobre no sería considerado como un error, pero el uso incorrecto de la terminología sí sería un error grave.

Una pregunta frecuente es cómo se relaciona MQM con el sistema de TAUS *Dynamic Quality Framework* (DQF) que describimos en la página 149. Ambos fueron desarrollados en más o menos el mismo período de tiempo, puede parecer en un primer momento que MQM y DQF estén en conflicto. Sin embargo, desde finales de 2014, las dos iniciativas están en contacto y trabajando para armonizar sus esfuerzos. Un examen detallado revela que sean ampliamente complementarias. MQM proporciona una manera de describir métricas arbitrarias de manera normalizada, pero no proporciona orientación sobre la interpretación de los resultados. DQF por el contrario, no tiene por objeto describir todos los posibles indicadores de calidad de traducción, pero proporciona orientación sobre la interpretación de las evaluaciones de calidad para escenarios específicos. Parte de la labor en curso es unificar los dos estándares juntos de tal manera que cualquier métrica DQF se puede describir en MQM. Esta armonización se plantea como un aspecto clave en un próximo proyecto financiado por la Unión Europea.

Para nuestra tesis no utilizaremos todos los elementos del marco; para nuestro trabajo es en primer lugar importante la tipología de errores (*issue types*) y los mecanismos de evaluación. No utilizaremos los mecanismos de puntuación propuestos ni crearemos una métrica propia. Es decir, solo utilizaremos las definiciones de errores de MQM, sin cambiar su estructura, ni añadir nuevos tipos de errores.

Definición de calidad y de error

Antes de describir detalladamente la clasificación de errores de MQM debemos indicar cómo se define en este contexto una traducción de calidad y cuál es la definición de error con que trabaja MQM. Una definición universalmente válida de calidad no existe, ya que la calidad es vista como algo relativo, que depende de varios factores como el propósito, la audiencia etc. Es por ello que, por ejemplo, en el proyecto QTLauch Pad (Uszkoreit 2014), que trabaja en el desarrollo de MQM, se habla de una definición multidimensional de calidad. Según esta perspectiva una traducción de calidad es aquella que demuestra la precisión y fluidez requerida para la audiencia y el propósito específico y que cumple con todas las especifici-

2. Tecnologías de la traducción, traducción automática y sus problemas

caciones negociadas anteriormente, teniendo en cuenta las necesidades del usuario final (Lommel, Uszkoreit, y Burchardt 2014).

Siguiendo esta definición un error representa cualquier problema que se pueda encontrar en el texto traducido que se presenta, o bien porque no corresponde con el original o porque se considera incorrecto en el idioma meta. MQM proporciona una tipología flexible de errores (*issue types*) que puede aplicarse a tareas de evaluación analítica de la calidad de la traducción. La lista de errores de MQM se divide en categorías y tenemos que anotar que la clasificación y jerarquización de errores de MQM puede variar dependiendo de la versión. La tipología que utilizaremos es la presentada en la versión de diciembre del 2015.

Categorías de errores de MQM

En la versión actual de MQM se proponen nueve categorías generales que agrupan sesenta y cinco tipos de errores, que representan los problemas más comunes que se encuentran en la evaluación de la calidad de los textos. Si entramos en detalle en cada categoría podemos ver que la definición agrupa más de 100 tipos de errores, que pueden ser o no seleccionados para la evolución en un proyecto concreto, lo que permite establecer según el proyecto un foco y un nivel mayor o menor de granularidad en la clasificación de los errores.

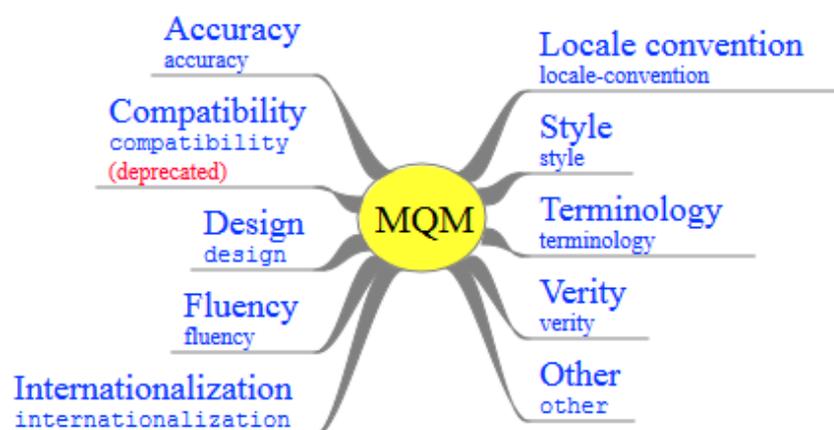


Ilustración 17: Categorías de errores de MQM (Lommel, Burchardt, Görög, Uszkoreit & Melby, 2016)

A continuación daremos una descripción general de cada una de las categorías que hacen parte de MQM, con las definiciones respectivas aportadas por los investigadores del *Deutsche Forschungszentrum für Künstliche Intelligenz* (DFKI) y del grupo de investigación

2. Tecnologías de la traducción, traducción automática y sus problemas

QTLaunchPad (Lommel et al. 2016). Las definiciones se encuentran en el documento original en inglés y las hemos traducido para su presentación en la tesis.

1) Accuracy

Bajo esta categoría encontramos los errores que tienen que ver con la exactitud misma de la traducción. Es decir con qué exactitud refleja el texto meta el texto original, teniendo en cuenta las diferencias autorizadas por las especificaciones.

En la gráfica siguiente los autores ilustran la relación jerárquica existente entre los tipos de errores que se agrupan en esta categoría:

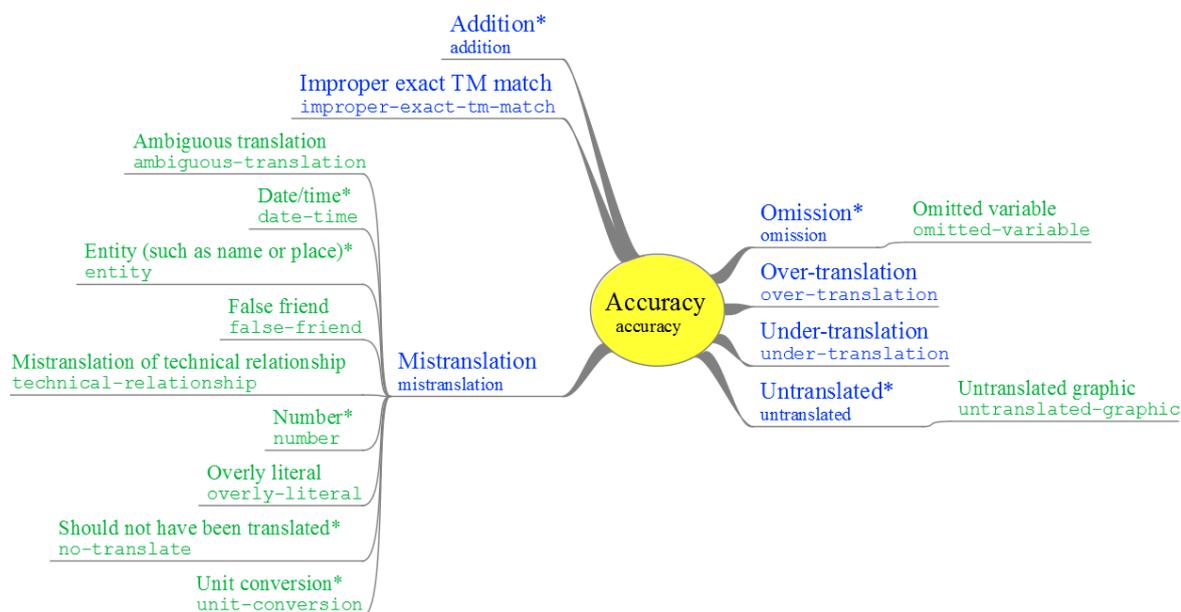


Ilustración 18: Jerarquía de los errores en la categoría Accuracy (Lommel et al. 2016)

Un error de la categoría *Accuracy* se presenta cuando el texto meta no refleja con precisión el texto original, permitiendo una diferencia no autorizada en las especificaciones del proyecto. En esta categoría nos encontramos con las siguientes subcategorías, que en algunos casos pueden tener subcategorías:

Categoría	Subcategoría	Definición
<i>Omission</i>		Un contenido que está presente en el original no se encuentra en la traducción.
	<i>Omitted variable</i>	Una etiqueta de posición variable se omite en la traducción.

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
<i>Over translation</i>		El texto traducido es más específico que el texto meta. El texto original se refiere a 'boy', pero se traduce con una palabra que se aplica solo a 'young boy', correcto sería utilizar el término general.
<i>Under-translation</i>		El texto meta es menos específico que el texto original. El texto original utiliza palabras que se refieren a un tipo específico de rango militar pero el texto meta se refiere a los militares en general.
<i>Untranslated</i>		Un contenido que debería haber sido traducido se ha dejado sin traducir.
	<i>Untranslated graphic</i>	El texto en un gráfico fue dejado sin traducir.
<i>Addition</i>		En una traducción se incluye(n) palabra(s) que no se encuentra(n) en el original.
<i>Improper exact match</i>		La traducción es una traducción exacta inadecuada, que ha sido extraída de una memoria de traducción.
<i>Mistranslation</i>		El contenido de la traducción no representa con precisión el contenido del original.
	<i>Ambiguous translation</i>	Un texto de original claro se traduce ambiguamente.
	<i>Date/time</i>	Fechas u horas no coinciden en el original y en la traducción.
	<i>Entity</i>	Nombres, lugares u otras “entidades” mencionadas no coinciden. P.ej. en el original se hace referencia a Dublín, Ohio y en la traducción a Dublín, Irlanda.
	<i>False friend</i>	La traducción ha utilizado incorrectamente una palabra que es aparentemente similar a la palabra original.
	<i>Mistranslation of technical relationship</i>	No se traduce con precisión un contenido técnico (aunque la traducción pueda parecer plausible).
	<i>Number</i>	No coinciden los números en el original y en la traducción.
	<i>Overly literal</i>	La traducción es excesivamente literal.
	<i>Should not have been translated</i>	Se traduce una parte del texto, una frase o una palabra que no debería haber sido traducida.
	<i>Unit conversion</i>	En la traducción no se han convertido valores numéricos para ajustarlos a las diferentes unidades (p.ej. monedas o métricas europeas frente los sistemas de medición de Estados Unidos).

Tabla 90: Categorías de errores que tienen que ver con *Accuracy*

2. Tecnologías de la traducción, traducción automática y sus problemas

2) Design

En esta categoría encontramos los errores relacionados con el diseño del contenido; los tipos de errores que se agrupan bajo esta categoría no tienen que ver con los aspectos lingüísticos de los textos y se pueden encontrar tanto en los textos originales como en sus traducciones.

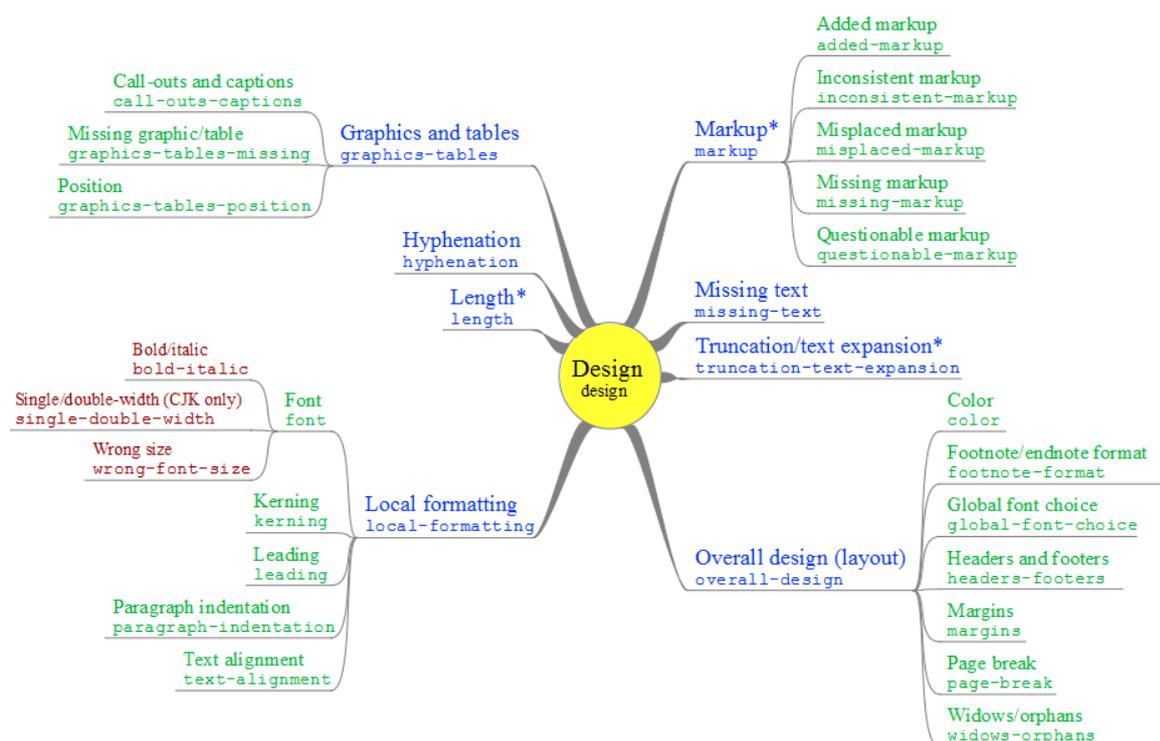


Ilustración 19: Jerarquía de los errores en la categoría Design (Lommel et al. 2016)

En la siguiente tabla presentamos todas las subcategorías que hacen parte de esta macrocategoría. Ya que la especificación de esta categoría es muy detallada y los nombres mismos son suficientemente explicativos, solo daremos la definición de las subcategorías y no la de las que se derivan de estas. En estos casos solo indicaremos los nombres en la columna de la subcategoría.

Categoría	Subcategoría	Definición
Graphics and tables	Call-outs and captions, missing graphic/table, position of graphic/table	Errores relacionados con el formato de gráficos y/o tablas.
Hyphenation		La separación de las palabras no es correcta.

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
<i>Length</i>		Existe una discrepancia significativa entre la extensión del texto original y su traducción.
<i>Local formatting</i>	<i>Font, Leading, Kerning, Paragraph indentation, Text alignment</i>	Errores relacionados con el formato del texto que se pueden encontrar tanto al texto original como al texto meta.
<i>Markup</i>	<i>Added markup; inconsistent markup; misplaced markup; missing markup; questionable markup</i>	Errores relacionados con las etiquetas.
<i>Missing text</i>		Hace falta parte del texto en la versión final.
<i>Truncation/text expansion</i>		El texto traducido no ocupa el espacio determinado para la visualización que se indicó en las especificaciones.
<i>Overall design (layout)</i>	<i>color; footnote/endnote format; global font choice; headers and footers; margins; page breaks; widows/orphans</i>	Problemas generales de diseño como color, tipo de letra, márgenes, etc.

Tabla 91: Categorías de errores que tienen que ver con *Design*

3) *Fluency*

Bajo esta categoría se agrupan los problemas o errores relacionados con la forma o el contenido de un texto, independientemente de si se trata de una traducción o no. Si un problema solo se puede detectar mediante la comparación del original y su traducción, no debe ser categorizado como error de fluidez en general, ni tampoco como algún tipo de las subcategorías que se describirán a continuación (DFKI GmbH y QTLaunchPad 2015). Este matiz de MQM es interesante porque resalta el hecho de que las categorías de errores propuestas no solo sirven para evaluar la calidad de traducciones. Sin embargo este aspecto no es relevante para nuestra tesis. La gráfica a continuación también hace parte de la documentación de la definición de errores de MQM citada anteriormente.

2. Tecnologías de la traducción, traducción automática y sus problemas

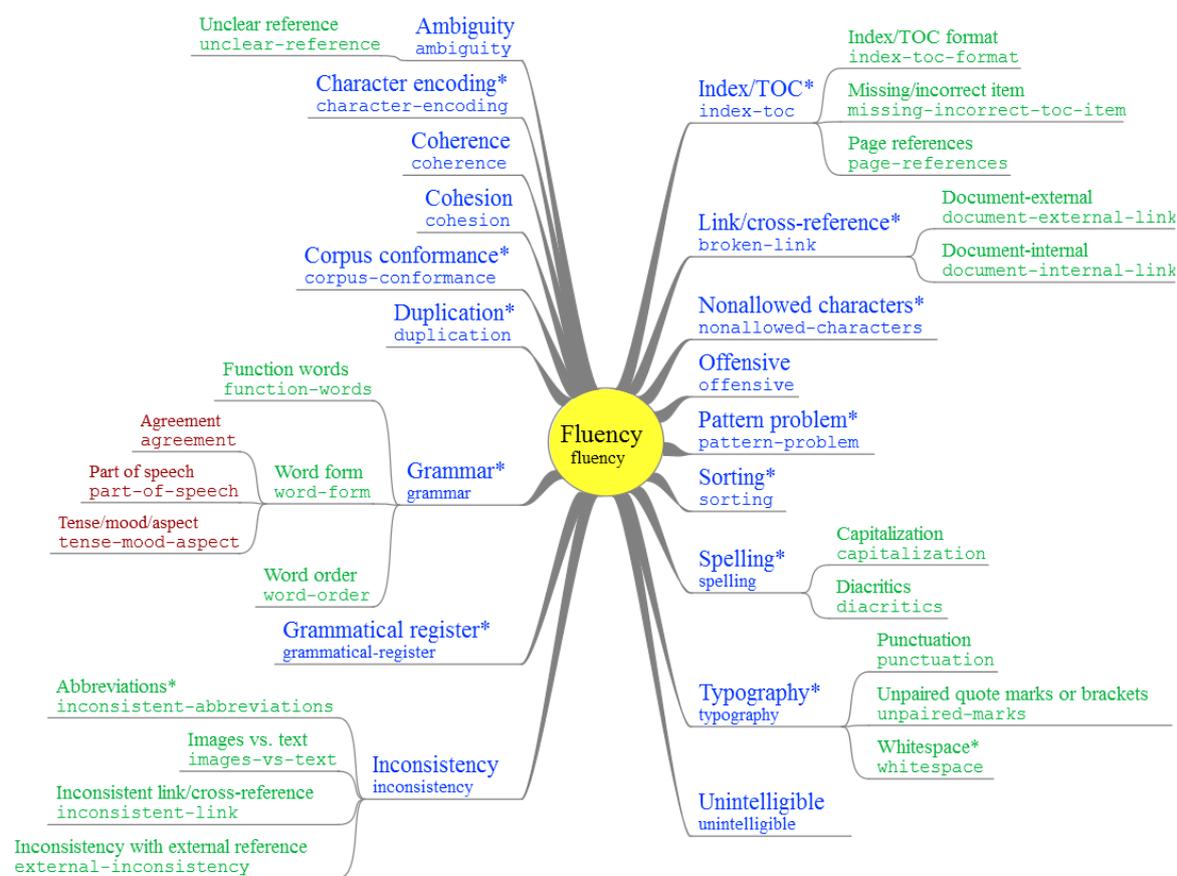


Ilustración 20: Jerarquía de los tipos de errores en la categoría Fluency (Lommel et al. 2016)

Del mismo modo que lo hicimos con la categoría *Accuracy* a continuación listaremos todas las subcategorías de errores que tienen que ver con la fluidez con sus respectivas definiciones.

Categoría	Subcategoría	Definición
<i>Ambiguity</i>		Un texto es ambiguo.
	<i>Unclear reference</i>	El texto utiliza pronombres relativos u otros mecanismos referenciales cuya referencia no está clara.
<i>Character encoding</i>		Los caracteres están distorsionados debido, por ejemplo, a la aplicación incorrecta de una codificación (UTF-8, ASCII, ISO Latin1, etc.).
<i>Coherence</i>		El texto en conjunto no tiene coherencia.
<i>Cohesion</i>		Faltan ciertas partes del texto (o no se emplean correctamente) que son necesarias para hacerlo comprensible; por ejemplo faltan referencias, conjunciones o cohesión léxica.

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
<i>Corpus conformance</i>		El contenido no tiene conformidad con el corpus de referencia.
<i>Duplication</i>		El contenido se ha duplicado erróneamente.
<i>Grammar</i>		Errores que tienen que ver con la gramática o la sintaxis de un texto, que no tienen que ver con la puntuación o la ortografía.
	<i>Function words</i>	Uso incorrecto de palabras preposiciones, verbos auxiliares, artículos o pronombres.
	<i>Word form</i>	La palabra no es correcta, por ejemplo cuando se construye erróneamente el pasado del verbo <i>to become</i> como <i>becomed</i> .
	<i>Agreement</i>	Dos o más palabras no están de acuerdo con respecto al caso, número, persona, u otras características gramaticales
	<i>Part of speech</i>	Una palabra equivocada en una oración, como cuando se utiliza erróneamente un adjetivo en lugar de un adverbio.
	<i>Tense/mood/aspect</i>	Flexión verbal incorrecta.
	<i>Word order</i>	Orden gramatical erróneo.
	<i>Grammatical register</i>	El texto tiene un nivel de formalidad más alto o más bajo que el requerido por las especificaciones o las convenciones generales del idioma.
<i>Inconsistency</i>		El texto muestra inconsistencia interna.
	<i>Abbreviations</i>	Uso inconsistente de las abreviaturas en el texto.
	<i>Image vs. text</i>	Una frase o palabra no es compatible con el texto que se muestra en las imágenes y/o en el texto escrito.
	<i>Inconsistent link/cross-reference</i>	Los enlaces son inconsistentes en el texto.
	<i>Inconsistency with external reference</i>	El texto es incompatible con una referencia externa especificada.
<i>Index/TOC</i>	<i>Index/TOC format; missing/incorrect item; page references</i>	Errores que tienen que ver con índices o contenidos, incluyendo el formato, las referencias de páginas, etc.
<i>Inconsistent link/cross-reference</i>		Los enlaces son inconsistentes en el texto.
	<i>Document-internal; document-external</i>	Enlaces o referencias incorrectas o inexistentes dentro del mismo documento o referencias externas.

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
<i>Non allowed characters</i>		El texto incluye caracteres que no son permitidos.
<i>Offensive</i>		El contenido es ofensivo de acuerdo con las especificaciones culturales o regionales.
<i>Pattern problem</i>		El texto contiene un patrón que no está permitido.
<i>Sorting</i>		Una lista no está ordenada correctamente.
<i>Spelling</i>		
	<i>Capitalization</i>	Uso errado de mayúsculas y minúsculas.
	<i>Diacritics</i>	Uso erróneo de signos diacríticos
<i>Typography</i>		Las cuestiones relacionadas con la presentación mecánica de texto. Esta categoría debe ser utilizada para cualquier error tipográfico distinto de la ortografía.
	<i>Punctuation</i>	Puntuación incorrecta.
	<i>Unpaired quote marks or brackets</i>	Uso erróneo o incorrecto de comillas o corchetes.
	<i>Whitespace</i>	Uso incorrecto de los espacios en blanco.
	<i>Unintelligible</i>	La naturaleza exacta del error no se puede determinar, pero implica una ruptura importante en la fluidez.

Tabla 92: Tipos de errores que tienen que ver con *Fluency*

4) *Verity*

Como mencionamos en la descripción inicial de esta métrica, MQM introduce el concepto de *Verity* en la evaluación de las traducciones (véase página 152). En esta categoría se agrupan los errores que se dan cuando en el texto se hacen declaraciones que contradicen el mundo objetivo del texto mismo. Por ejemplo cuando se hace referencia a una función que está disponible en un determinado modelo de un automóvil, pero en realidad dicha función no está disponible en el modelo en cuestión. A continuación veremos la imagen que ilustra la organización jerarquía de los tipos de errores que encontramos en esta categoría.

2. Tecnologías de la traducción, traducción automática y sus problemas

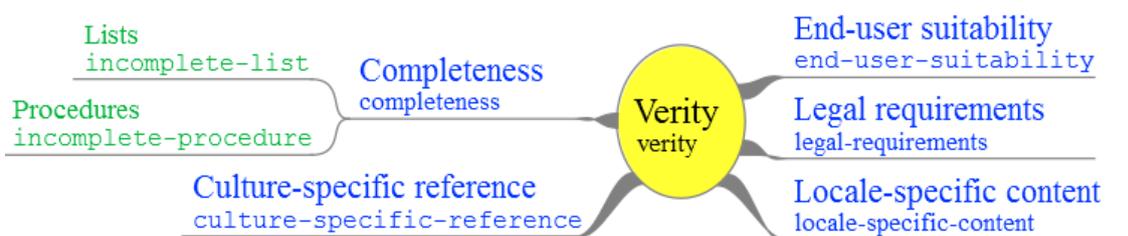


Ilustración 21: Jerarquía de la categoría de errores *Verity* (Lommel et al. 2016)

Como podemos ver la categoría *Verity* tiene 5 subcategorías principales, cuyas definiciones damos en la siguiente tabla.

Categoría	Subcategoría	Definición
<i>End-user suitability</i>		El contenido no es adecuado para la utilización por parte del usuario final, con exclusión de los problemas relacionados con la idoneidad para el entorno local de destino
<i>Legal requirements</i>		Un texto no cumple los requisitos legales que se indican en las especificaciones
<i>Locale-specific content</i>		Contenido específico de la localidad de origen no se aplica a la localidad de destino, el público o el propósito que se persigue.
<i>Completeness</i>		El texto no está completo.
	<i>Incomplete List</i>	Faltan algunos ítems en la lista.
	<i>Incomplete procedure</i>	En un procedimiento faltan pasos necesarios.
<i>Culture-specific reference</i>		Se utiliza inapropiadamente una referencia específica cultural que no va a ser comprendida por la audiencia. Un texto de marketing en griego incluye una referencia a la música popular griega. Cuando se traduce al inglés esta referencia no tendrá sentido.

Tabla 93: Tipos de errores de la categoría *Verity*

Aunque los errores de la categoría *Verity* pueden aparecer tanto en el texto original como en el texto meta, a menudo surgen durante la traducción cuando, por ejemplo, una declaración que es válida para la localidad o el contexto original, no es válida en la localidad o contexto meta.

2. Tecnologías de la traducción, traducción automática y sus problemas

5) Terminology

En la versión actual de la clasificación de tipos de errores de MQM se separó esta categoría y algunas subcategorías se encontraban en las otras versiones en la categoría *Accuracy* o en la categoría *Fluency* se incluyeron ahora en la categoría *Terminology*. En el apartado 2 del documento en que se describen las categorías de errores se encuentra un listado de las diferencias o modificaciones realizadas con respecto a las versiones anteriores de MQM (Lommel et al. 2016).

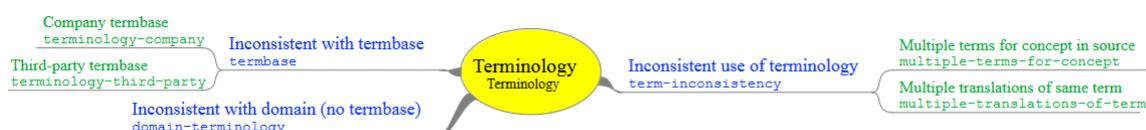


Ilustración 22: Jerarquía de la categoría de errores *Terminology* (Lommel et al. 2016)

Un error de terminología se da cuando un término (palabra de dominio específico) se traduce con un término que no está previsto para el dominio o que especifique lo contrario. Los editores recomiendan no utilizar esta categoría para marcar una palabra que se ha traducido correctamente, pero no con el sentido correspondiente en el contexto. Por ejemplo en la combinación *river bank* se traduce *bank* al castellano como la institución financiera y no en el sentido que tienen al estar combinada con el sustantivo *river*, en este caso no tendríamos un problema de terminología sino un caso normal de *mistranslation*.

Categoría	Subcategoría	Definición
<i>Inconsistent with termbase</i>	<i>Company terminology</i>	El texto infringe las directrices de terminología de la empresa/organización indicada en la base de datos terminológicos correspondiente.
	<i>Third-party termbase</i>	
<i>Inconsistent with domain (no termbase)</i>		Un término no se usa conforme a las expectativas generales de dominio. Un texto financiero se utiliza <i>deduct</i> en lugar de <i>debit</i> . Aunque conceptualmente podrían ser sinónimos, <i>deduct</i> está violando convenios de dominio.
<i>Inconsistent use of terminology</i>		No hay consistencia en la utilización de la terminología en todo el texto.
	<i>Multiple terms for concept in source</i>	

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
	<i>Multiple translations of same term</i>	

Tabla 94: Tipos de errores de la categoría *Terminology*

Errores de la categoría *Terminology* pueden encontrarse también en el texto original y no solo en la traducción.

6) *Style*

Bajo esta categoría encontramos los errores estilísticos. Errores de estilo se dan también cuando, por ejemplo, encontramos en la traducción de una campaña de publicidad que debe tener un estilo alegre y liviano, un estilo serio y pesado, aunque en las especificaciones se indicó que se debía mantener el estilo del original.

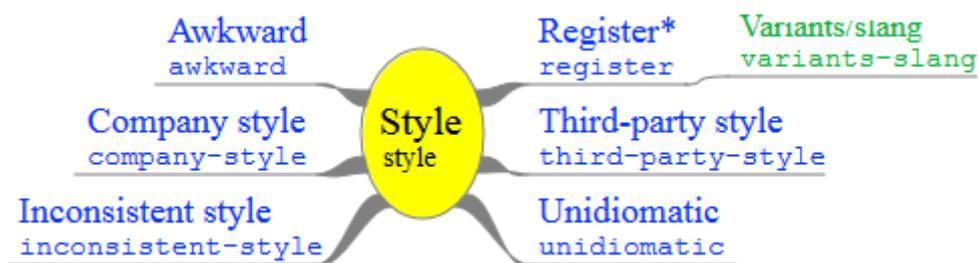


Ilustración 23: Jerarquía de la categoría de errores *Style* (Lommel et al. 2016)

En la categoría *Style* encontramos seis subcategorías. Mientras que la subcategoría *Awkward*, se aplica solo a los errores encontrados en las traducciones, las otras subcategorías pueden ser utilizadas para marcar los errores tanto en los textos originales como en los textos meta. Veamos a continuación la definición de cada una de las subcategorías.

Categoría	Subcategoría	Definición
<i>Register</i>		El texto utiliza un nivel de formalidad alta o más baja que la requerida por las especificaciones o las convenciones del lenguaje en general.
	<i>Variants/slang</i>	El texto utiliza palabras informales que no son apropiados para el registro previsto.

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
<i>Third-party-style</i>		El texto viola una guía de estilo de terceros, por ejemplo si se deben utilizar guías de estilo como la APA.
<i>Unidiomatic</i>		El contenido es gramatical, pero no idiomático.
<i>Awkward</i>		Un texto está escrito en un lenguaje torpe.
<i>Company style</i>		El texto no es compatible con el estilo de la compañía.
<i>Inconsistent style</i>		No hay consistencia en la utilización del estilo en todo el texto.

Tabla 95: Tipos de errores de la categoría *Style*

7) *Locale Convention*

En esta categoría se agrupan los errores que hallamos cuando el texto no respeta las convenciones específicas regionales y viola los requisitos para la presentación de contenidos en la localidad meta. Errores de esta categoría se pueden dar tanto en el original como en la traducción. Aquí nos encontramos errores tales como la utilización errónea de un punto para separar miles en alemán, ya que en este idioma se utiliza para ello la coma.

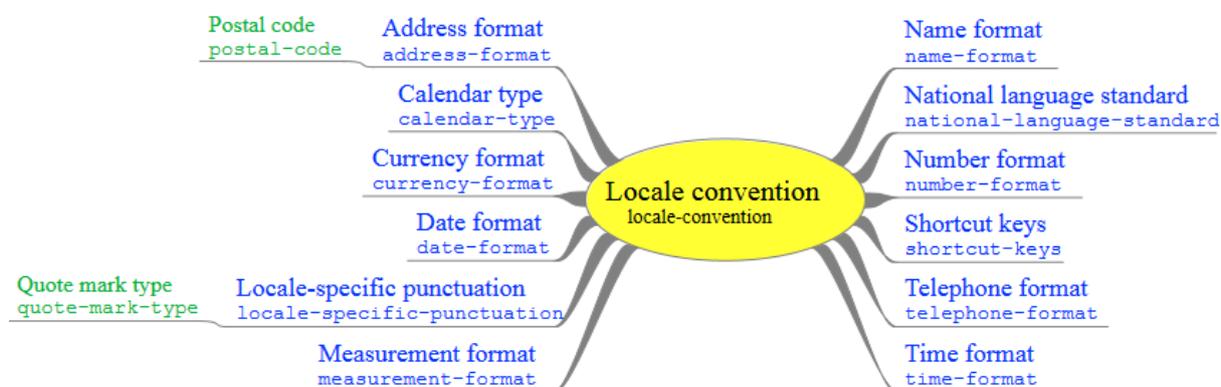


Ilustración 24: Jerarquía de la categoría de errores *Locale convention* (Lommel et al. 2016)

La categoría *Locale convention* tiene doce subcategorías que describen los tipos de errores que como se puede ver ya en la ilustración anterior no tienen que ver con las convenciones en sentido cultural, sino se trata de convenciones de índole mecánico o de tipos de formatos. Los errores de convenciones en sentido cultural han sido clasificados por MQM en la subcategoría *locale-specific-content* de la categoría *Verity* (véase página 161).

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
<i>Address format</i>		Se utiliza el formato incorrecto para las direcciones.
	<i>Post code</i>	Se utiliza la forma incorrecta de códigos postales ⁶ .
<i>Calendar type</i>		Se utiliza el de calendario erróneo, p.ej. el calendario islámico en lugar del gregoriano o viceversa.
<i>Currency format</i>		Se utiliza el formato incorrecto de moneda.
<i>Date format</i>		Uso incorrecto d el formato de fecha.
<i>Locale-specific punctuation</i>		El texto utiliza automáticamente puntuación que no es apropiada para el entorno local especificado, por ejemplo cuando un texto traducido del inglés al japonés mantiene al estilo europeo de puntuación.
	<i>Quote mark type</i>	Se utilizan las comillas inapropiadas, por ejemplo en un texto en francés debe utilizarse las comillas angulares «...».
<i>Measurement-format</i>		Se emplea un formato de medición inapropiado para la configuración regional.
<i>Name format</i>		Un texto viola las normas del idioma meta de representación de los nombres.
<i>National language standard</i>		Un texto viola las normas del idioma nacional, p.ej. la academia francesa prohíbe la utilización de anglicismos en textos impresos.
<i>Number format</i>		Se utiliza inapropiadamente el formato numérico.
<i>Shortcut keys</i>		Un producto de software traducido utiliza accesos directos que no se ajustan a las expectativas del entorno local y que no tienen sentido en dicho entorno.
<i>Telephone-format</i>		Empleo erróneo del formato para los números telefónicos.
<i>Time format</i>		Se utiliza un formato de hora inapropiada para su localización. Un texto escrito para los EE.UU. Se utiliza una notación de 24 horas en lugar de las horas seguidas de AM/PM
<i>Locale convention</i>		El texto no se ajusta a las convenciones específicas de la región y viola los requisitos para la presentación de contenidos en la localidad de destino.

Tabla 96: Tipos de errores de la categoría *Locale convention*

⁶ Este problema o error se puede categorizar también como un error de internacionalización (*localization*)

2. Tecnologías de la traducción, traducción automática y sus problemas

8) *Internationalization*

Internacionalización cubre áreas relacionadas con la preparación del contenido original para su traducción o posterior localización. Problemas de internacionalización pueden ser detectados a través de los problemas que se encuentran en la traducción (en particular de los descritos en la categoría *locale convention*), pero una revisión de los errores de internacionalización se realiza generalmente por separado a partir de una evaluación general de la calidad de la traducción.

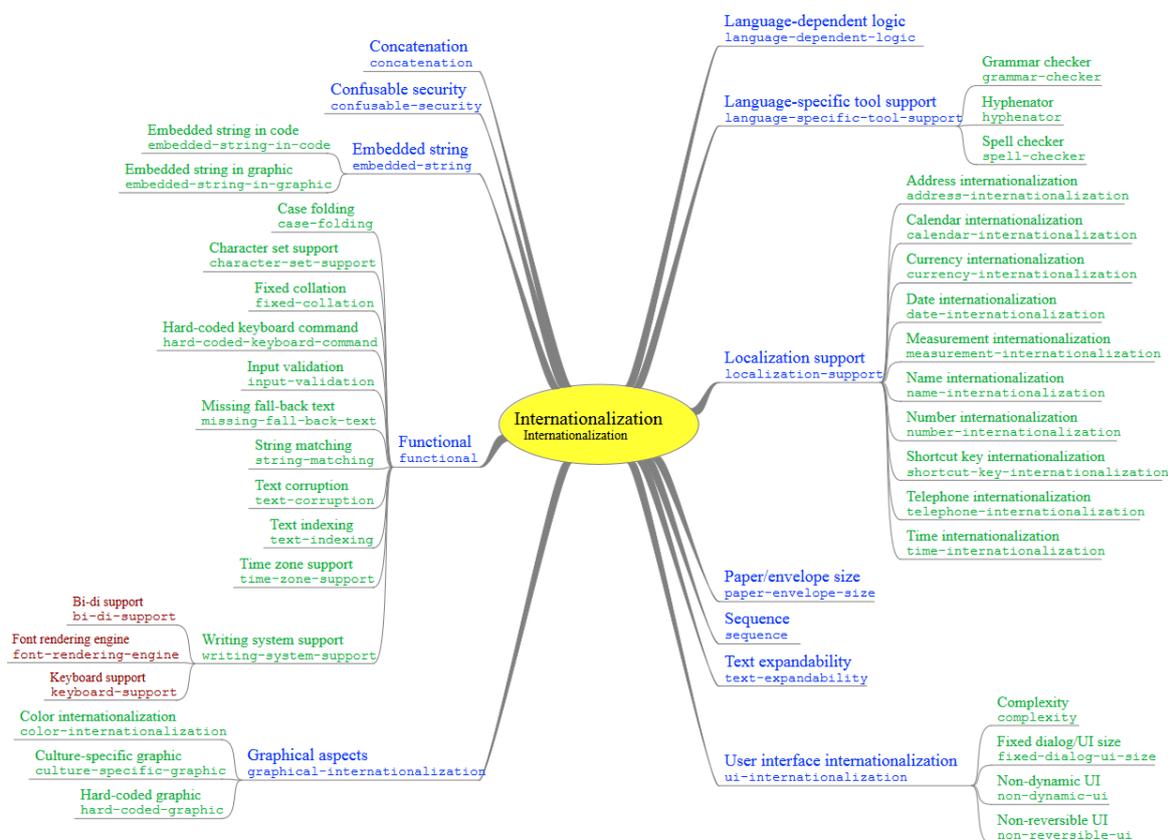


Ilustración 25: Jerarquía de la categoría de errores *Internationalization* (Lommel et al. 2016)

En la gráfica anterior podemos ver como la definición de MQM es bastante detallada en esta categoría, aquí solo daremos la definición general de las doce categorías principales, nombraremos las subcategorías, pero no detallaremos su definición, ya que la denominación misma de la subcategoría deja claro de que se trata:

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
<i>Concatenation</i>		El texto se enlaza de manera que no funciona correctamente cuando se traduce el texto. Este caso se da en un segmento como <i>You have found the "\$item". Do you wish to pick it up?</i> en el que no se puede reconocer el género de la variable " <i>\$item</i> " y por ello no se puede traducir correctamente el artículo <i>the</i> ni el equivalente de <i>it</i> .
<i>Confusable security</i>		El software no proporciona ninguna protección de seguridad contra caracteres fácilmente confundibles como escritura Latina A, griega A y cirílica A. Los usuarios pueden seleccionar nombres de usuario con los caracteres Unicode válidos, sin control.
<i>Embedded string</i>	<i>Embedded string in code; embedded string in graphic</i>	En el contenido textual se añaden otros elementos de manera que lo hacen inaccesible durante el proceso de localización.
<i>Funcional</i>	<i>Case folding; character set support; fixed-collation; hard-coded-keyboard-command; input-validation; missing-fall-back-text; string-matching; text-corruption; text-indexing; time-zone-support; writing-system-support</i>	Errores funcionales se dan por la falta de apoyo para los idiomas, scripts, guiones u otras características específicas del entorno local.
<i>Graphical aspects</i>	<i>Graphical-internationalization; color-internationalization; culture-specific-graphic; hard-coded-graphic</i>	Aspectos gráficos del contenido que no se pueden cambiar fácilmente para que coincidan con las expectativas específicas locales. El código del software contiene gráficos a los que no se puede acceder y por lo tanto no se pueden cambiar para ser localizados, ni cambiados durante el proceso de localización.
<i>Language-dependent logic</i>		El contenido incluye supuestos lógicos idioma o dependientes de la localidad que le impiden ser localizado apropiadamente
<i>Language-specific-tool-support</i>	<i>Grammar checker, hyphenator; spell checker</i>	Ciertas herramientas tienen la opción de determinar un idioma pero no brindan el apoyo para este.

2. Tecnologías de la traducción, traducción automática y sus problemas

Categoría	Subcategoría	Definición
<i>Localization support</i>	<i>Address, calendar, currency, date, measurement, name, number, shortcut key, telephone, time-internationalization</i>	Aspectos de tienen que ver con la manera en que un producto de software localiza (o no) datos sensibles. Por ejemplo un formulario en línea para hacer citas que se empleará en Europa, no permite incluir datos que no sean de los Estados Unidos.
<i>Paper/envelope size</i>		Los sistemas no son compatibles con el formato de papel o sobres utilizados en la localidad.
<i>Resource externalization</i>		Recursos traducibles no se han externalizado adecuadamente en un código funcional. Por ejemplo si en un aviso legal en alemán se utiliza <i>Du</i> o sea lenguaje familiar en lugar de <i>Sie</i> (estilo formal).
<i>Sequence</i>		Las secuencias de gráficos o texto aparecen en un orden específico de la cultura original que no tiene sentido en otros lugares. Por ejemplo si se traduce al árabe o al hebreo que utilizan escritura de derecha a izquierda.
<i>Text expandability</i>		No hay suficiente margen para permitir la expansión de texto
<i>User interface internationalization</i>	<i>Complexity; fixed dialog/UI size; non-dynamic UI; non-reversible UI</i>	Una interfaz de usuario no ha sido debidamente internacionalizada y así muestra problemas en las versiones localizadas, por ejemplo si se hacen separaciones en la versión original en inglés, que no tienen sentido cuando se traducen a otro idioma.

Tabla 97: Tipos de errores de la categoría *Internationalization*

En el marco metodológico indicaremos de qué manera utilizamos estas categorías y qué aportes tuvieron para nuestro análisis del corpus.

9) Other

Esta dimensión se utiliza para cuestiones que no pueden ser clasificadas de otro modo en una dimensión de MQM, según los autores en la práctica *Other* solo se utiliza en muy raras ocasiones.

2. Tecnologías de la traducción, traducción automática y sus problemas

2.2.5.3.3 *Appraise*

Bajo este nombre se agrupa un kit de herramientas de código abierto para la evaluación manual de salidas de traducción automática (Federmann 2010). *Appraise* permite recoger juicios humanos sobre las salidas de traducción y además implementar de tareas de anotación como 1) el control de calidad de traducción, 2) la clasificación de las traducciones, 3) de clasificación de errores y 4) finalmente la posesición manual. Se utiliza un formato extensible para importar/exportar datos y se puede adaptar fácilmente a las nuevas tareas de anotación. Las tareas de anotación se explican con más detalle en el documento. La versión actual también incluye el cálculo automático de las puntuaciones y se puede acceder fácilmente a los resultados de la evaluación. Esta herramienta que se ha utilizado ya en varios de proyectos de investigación comenzó a desarrollarse en 2009 como parte del proyecto EuroMatrixPlus en el que se utilizó como herramienta para comparar rápidamente diferentes propuestas de traducción automática para un proyecto del DFKI con el fin de evaluar los avances en el desarrollo de un motor híbrido (GitHub 2014). El software utiliza XML como formato importar y exportar que puede ser adaptado a otros formatos, además permite realizar las siguientes tareas:

1. Clasificación: el evaluador puede ver el segmento original y varias propuestas de traducción. También puede visualizarse —si se desea— una traducción de referencia para apoyar al evaluador en el proceso de clasificación. Se implementó además un modelo de clasificación con tres alternativas para cada par de oraciones (original y traducción) para facilitar el proceso $A > B$, $A=B$, o $A < B$.
2. Estimación de calidad: el evaluador puede clasificar las propuestas de traducción como aceptables, puede ser mejorada o ninguna de las dos. Aquí también se tiene la opción de ver, si está disponible, la traducción de referencia.
3. Clasificación de error: el evaluador tiene aquí la posibilidad definir cada traducción a nivel de oración o de palabras. También puede saltarse la evaluación indicando que la propuesta tiene muchos errores: en caso de que no se pueda diferenciar entre errores menores y graves.
4. Posedición: aquí se tiene la posibilidad de corregir la frase que fue calificada anteriormente como ‘fácil de poseer’.

2. Tecnologías de la traducción, traducción automática y sus problemas

Appraise fue utilizado no solo durante el proyecto EuroMatrixPlus, sino también para la evaluación del proyecto taraXÚ, T4ME, también en el proyecto ML4HMT para la evaluación de las propuestas de traducción. Finalmente se utilizó en el contexto traducción de terminología para el proyecto Monnet del ámbito financiero.

2.3 Conclusiones del capítulo

Las tecnologías de la traducción y en particular la traducción automática avanzan continuamente y cada vez parece arrojar mejores resultados. Los motores de traducción automática están elaborados para establecer la mejor concordancia entre el original y la traducción.

Después de este resumen en parte diacrónico de los principales métodos utilizados hasta el momento para la traducción automática de textos podemos afirmar que no se ha llegado aún a alcanzar un resultado como el de la traducción humana, pero que se están desarrollando métodos y motores, en especial híbridos, que combinan las ventajas de los sistemas desarrollados hasta ahora.

Los sistemas basados en reglas utilizan para el análisis morfológico de la lengua original representaciones explícitas de información lingüística como diccionarios monolingües, reglas para la desambiguación léxica, gramáticas para el análisis sintáctico, reglas para la transferencia estructural y/o diccionarios monolingües para la generación morfológica en el idioma meta. En cambio los sistemas estadísticos —que aprenden a partir de una gran cantidad de corpus paralelos y monolingües— no necesitan de ninguna información lingüística, lo que permite que puedan funcionar casi sin intervención humana.

Las traducciones de los sistemas basados en reglas son por ello mismo más predecibles: tanto los aciertos como los errores pueden repetirse, ya que son el resultado de la aplicación de dichas reglas. Las traducciones de los sistemas basados en corpus y estadísticas pueden ser más fluidas, pero menos fieles. Esto hace plausible el desarrollo de sistemas híbridos que combinen las ventajas de cada sistema y minimicen sus deficiencias.

Después de clasificar los diversos problemas y errores que encontramos en la traducción automática podemos intuir que la ambigüedad —tanto semántica como estructural o sintáctica— constituye el problema u obstáculo primordial para la traducción automática. Este

2. Tecnologías de la traducción, traducción automática y sus problemas

tipo de problema y los demás problemas y errores mencionados que se presentan en la traducción automática no pueden ser resueltos sin la intervención humana, es decir, sin ayuda del proceso de posesición. El proceso de posesición es necesario e indispensable si se requiere alcanzar un nivel alto de exactitud y fluidez en las traducciones. Es por ello que en el siguiente capítulo nos dedicaremos al tema de la posesición.

3. Posedición

Como vimos en el capítulo anterior por más adelantos que se hayan dado en el campo de la traducción automática, el desarrollo actual de los sistemas de traducción automática requiere de la participación de seres humanos para poseer y así garantizar la calidad de las traducciones (Specia y Farzindar 2010). En el presente capítulo partiremos de la definición del concepto y tarea de posedición. También mencionaremos los métodos de control, las herramientas disponibles y mostraremos algunas guías de posedición ya existentes.

3.1 Definición

Existen diferentes definiciones de lo que es posedición que van desde la más general en la que se entiende como un proceso de “modification rather than revision” (Loffler-Laurian 1984), pasando por una definición que se centra en la persona que realiza la tarea “term used for the correction of machine translations output by human linguists/editors” (Veale y Way 1997) y otras que describen más el procedimiento mismo “checking, proof-reading and revising translations carried out by any kind of translating automation” (Gouadec 2007). Allen considera que la posedición es “by far most commonly associated as a task related to MT” (Allen 2003, 297). Según TAUS posedición es “*the correction of machine-generated translation output to ensure it meets a level of quality negotiated in advance between client and post-editor*”(TAUS/CNGL 2011). También existe la idea generalizada de que la posedición solo implica un mínimo de trabajo, mientras que otros autores le otorgan incluso la función de ‘reparar textos’ (Klings 2001). En nuestra tesis hablaremos de posedición como el paso o conjunto de pasos en un proceso de traducción ya iniciado, en el que un poseedor modifica, edita y/o corrige los textos que han sido traducidos con ayuda de un motor de traducción automática.

El proceso de posedición es diferente del proceso normal de corrección de un texto traducido convencionalmente. Primero, porque el tipo de errores que se encuentran es diferente. Segundo porque si se encuentra en una traducción automática una ‘interpretación errónea’ del significado, estamos hablando de niveles diferentes, ya que en la traducción automática no se realiza una ‘interpretación’ en sentido propio, sino solo se aplican reglas o formulas estadísticas; es decir que si el significado no se reproduce correctamente se debe a que la aplicación de estas reglas no funcionó adecuadamente. En tercer lugar debido a que con la posedi-

3. Posedición

ción se pretende, como vimos en la definición de TAUS, obtener un buen resultado en un mínimo de pasos. Y por último porque en la corrección de traducciones se pretende dar desde un comienzo una calidad final óptima; en el caso de la posedición la calidad puede ser diferente, dependiendo —como veremos más adelante— del tipo de posedición que se realice (McElhane y Vasconcellos 1988). La tarea del poseedor consiste primero en determinar si el segmento traducido cumple con ciertos requisitos mínimos y si puede ser corregido en un lapso tiempo razonable; en segundo lugar en editar el segmento traducido y por último en traducir los segmentos para los que la traducción automática no ha arrojado ningún resultado válido.

El que el proceso de posedición sea rentable (o no), es decir, sea más efectivo que realizar una nueva traducción, depende de tres factores principales (Elsen 2012) en primer lugar de la duración de la evaluación en que se decide si se puede utilizar o no el segmento traducido. Por esta razón se habla a menudo de la denominada “*2 second rule*”, según la cual la evaluación no debe durar más de dos segundos. Si durante este lapso de tiempo de dos segundos no se ha terminado con la evaluación, no se debe editar la traducción automática. Es importante establecer ciertas líneas que determinen hasta qué punto una traducción automática es aceptable o no. El poseedor debe conocer en primer lugar qué fin tiene la traducción que debe editar y qué calidad debe tener la traducción, es decir si la traducción va a ser por ejemplo publicada, o si solo es necesario tener una idea general del contenido del texto traducida.

Qué la posedición sea rentable depende en segundo lugar del esfuerzo requerido para retocar los segmentos utilizados. Los segmentos traducidos automáticamente deben tener tan buena calidad que se requiere poco para su corrección o ajuste, es decir, la traducción automática debe ser por lo menos tan buena como un *Fuzzy Match*. Siempre hay que tener en cuenta que el esfuerzo de corrección —como anunciamos anteriormente— depende de la calidad que se espera. Como describiremos en el capítulo “Niveles de posedición” a partir de la página 176, existen diferentes niveles en el trabajo de posedición, que van desde el más elemental y mínimo, hasta el que tiene como objetivo producir un texto de alta calidad.

En tercer lugar el que el proceso de posedición sea rentable también depende del número de segmentos traducidos que se puedan utilizar. La cantidad de segmentos que se pue-

3. Posedición

den utilizar debe ser considerable, sino el proceso no será productivo y sería más fácil hacer una nueva traducción.

3.2 *Tipos de posedición*

Existen diferentes tipos de posedición. Siguiendo la clasificación de Tatsumi (2010, 11–14) podemos hablar de 1) la posedición realizada por personas y 2) la posedición ayudada o sistematizada. Según Tatsumi la *Human posedition* es realizada por los poseedores con el fin de producir rápidamente, con el mínimo de cambios posibles, textos aceptables y con cierto grado razonable de precisión. Este tipo de posedición requiere mucho esfuerzo, ya que es un proceso repetitivo y tedioso, que por ello mismo no está exento de errores (Wagner 1985, 203). Para evitar este tipo de problemas se han desarrollado programas o herramientas que pretenden minimizar la tarea de posedición y fomentar el desarrollo de entornos interactivos de posedición (Tatsumi 2010, 12). De este tipo de iniciativas surgieron programas como *Reverso*, desarrollado por la compañía Softissimo («Reverso» 2016), que combinaban de manera interactiva traducción automática y la posedición con el fin de disminuir el tiempo requerido para la posedición (Allen 2001, 27). Encontramos otro ejemplo de este tipo de ayudas para la traducción en el programa *Caitra* (Koehn 2009) que proporciona tres tipos de herramientas que permiten preeditar los textos (completando frases), traducción automática (con ayuda de un motor basado en estadísticas) y posedición; según el estudio de Koehn todos estos tipos de asistencia ayudaron a aumentar la velocidad y la calidad de la traducción en comparación con la traducción humana no asistida.

En el momento existe la tendencia de utilizar los recurso de traducción, con los que los traductores ya están familiarizados, para implementar las herramientas de posedición (O'Brien 2010). Flórez (2012, 221–229) presenta un catálogo de herramientas para la posedición que divide en tres secciones 1) aquellas que están vinculadas a entornos de traducción 2) aquellas que hacen parte de motores de traducción automática y 3) herramientas complementarias a la tarea de posedición. Pero en nuestra tesis no nos ocuparemos de las herramientas que se pueden emplear para la posedición, sino de la tarea misma. Por ello continuaremos nuestra presentación con los niveles de la posedición.

3. Posedición

3.3 Niveles de posesición

La posesición puede realizarse de varias maneras y por ello podemos hablar de diversos niveles de posesición, que se diferencian de acuerdo al propósito o la audiencia que se tiene (Allen 2003). Es decir, si la traducción posesitada está destinada a la publicación, o si solo se quiere tener una idea general del contenido de un documento. Si solo queremos saber de qué se trata el texto traducido hablamos de *Gisting* como nivel básico o de una posesición rápida (RPE). *Gisting* es lo que todos hacemos al utilizar un traductor automático en línea — por ejemplo el traductor en línea de Google Translate®.

A diferencia de *Gisting* en el proceso de posesición rápida o RPE un hablante de la lengua meta realiza la corrección. No se deben realizar sino cambios mínimos, como por ejemplo, el sustituir palabras desconocidas, eliminar redundancias o corregir ambigüedades que distorsionen el sentido. Los textos que sean corregidos con este método tampoco están destinados a la publicación y la calidad de la traducción es baja o media. A continuación damos un ejemplo, en el que marcamos además los errores y las correcciones, las marcaciones son nuestras.

Original	<i>They were added in MySQL 5.1.2</i>
Traducción automática	<i>Sie waren addded in MySQL 5.1.2</i>
Posedición	<i>Sie waren in MySQL 5.1.2 hinzugefügt</i>

Tabla 98: Ejemplo de RPE (Schumann y Sivaloganathan 2013)

Si el texto posesitado debe ser publicado hablamos de MPE (posesición mínima) o de posesición en el sentido convencional (CPE). MPE es utilizado especialmente en automoción para la corrección de traducciones de manuales de manejo y mantenimiento de maquinaria y ciertos textos que no se leen linealmente (listas, enumeraciones). Al posesitar deben eliminarse todo aquello que sea extraño o confuso. El resultado tiene una calidad media, que permite el que se pueda repartir a otras instancias. Como en el ejemplo anterior marcamos los errores y la corrección.

3. Posedición

Original	<i>Eventually the system was migrated to Linux</i>
Traducción automática	<i>Schließlich wurde das System zu Linux gewandert</i>
Posedición	<i>Schließlich wurde das System zu Linux migriert</i>

Tabla 99: Ejemplo de MPE (Schumann y Sivaloganathan 2013)

En el proceso de posedición convencional CPE el poseedor realiza los cambios que un traductor hubiese realizado, sin cambiar el estilo. Este tipo de posedición puede en algunas ocasiones demandar más trabajo que si se hubiese traducido el segmento original sin intervención de ningún motor de traducción automática ya que el resultado debe ser tan bueno, como si hubiese sido traducido por una persona.

Original	<i>The result file is created and its contents overwritten, even if an error occurs while generating the dump:</i>
Traducción automática	<i>Die Ergebnisdatei wird und sein Inhalt, überschreiben, auch wenn ein Fehler auftritt, eingerichtet, während er den Auszug generiert:</i>
Posedición	<i>Die Ergebnisdatei wird eingerichtet und ihr Inhalt überschrieben, auch wenn ein Fehler auftritt, während der Auszug generiert wird:</i>

Tabla 100: Ejemplo de CPE Traducción automática

En el tabla siguiente mostraremos también con ejemplos las diferencias entra posedición rápida y posedición completa (De Palma 2013).

Segmento original	<i>Erste Abgabe Dokumentation betriebsrelevante Ersatzteile</i>
Traducción automática	<i>first delivery documentation operation-relebant spare parts</i>
Light posedition	<i>First delivery documentation operation-relebant spare parts</i>
Full posedition	<i>Initial submission of documentation for operations-relevant replacement parts</i>

Tabla 101: Ejemplo comparativo de los tipos principales de posedición

Como podemos ver en la posedición rápida solo se ajustó el uso de mayúsculas y minúsculas. En la posedición completa se cambia la frase de tal manera que se transmita real-

3. Posedición

mente el sentido del original. TAUS (2015) nos brinda de nuevo una comparación sistemática de estos tipos principales de posesición:

Nivel	Posedición rápida	Posedición completa
Gramática	Se aceptan errores menores.	Debe ser correcta.
Puntuación	Se aceptan variaciones o errores.	Debe ser correcta y consistente.
Ortografía	Se aceptan variaciones o errores menores.	Debe ser consistente, incluso la separación de palabras.
Terminología	Debe ser comprensible y utilizable.	Debe ser precisa y consistente.
Estilo y tono	No debe ser ofensivo.	Debe adecuarse al contenido.
Estilo	Se aceptan variaciones.	Debe ser consistente (títulos, lista de artículos etc.).
Formato	No es importante.	Debe ser consistente con respecto al texto original

Tabla 102: Diferencias entre la posesición rápida y completa según TAUS

También nos encontramos aquí con una distinción importante. Podemos hablar de posesición en uno o dos idiomas (mono o bilingüe). En la posesición monolingüal el poseedor solo podrá ver el resultado de la traducción automática, pero no el segmento original. En este caso si la traducción es correcta pero no transmite el sentido del original el poseedor no podrá reconocer ni corregir el error.

3.4 Reglas de posesición

No solo existen niveles de posesición, también se han desarrollado reglas para el proceso mismo. Ya en 1985 enumeró Wagner (1985, 7) ciertas reglas a seguir para facilitar la posesición, aquí listamos algunos de estas reglas:

- (En lo posible) leer el original antes de empezar la corrección.
- Conservar lo más posible de la traducción automática.
- No detenerse mucho tiempo en un problema.

3. Posedición

- No ocuparse en cuestiones de estilo, es decir, si el estilo es adecuado para el texto en cuestión.
- No invertir tiempo en ningún tipo de investigación, solo usar ayudas o métodos de consulta rápidos (por ejemplo, diccionarios en línea o bases de terminología).
- Solo hacer cambios en caso de que sea absolutamente necesario. Por ejemplo, corregir palabras o frases que a) sean absurdas b) incorrectas y si aún hay tiempo) corregir ambigüedades.

McElhaney y Vasconcelos (1988) propusieron unas de las más simples reglas para la posedición de textos: 1) trabajar en el texto de izquierda a derecha; 2) evitar hacer cambios importantes en el orden de las frases y 3) hacer uso de ayudas mecánicas como el ratón o la función de 'búsqueda'. Estas simples reglas produjeron después de un mes de práctica que el rendimiento del trabajo de los poseedores aumentara. Ya desde entonces era claro que no es posible definir reglas universales para la posedición que reduzcan el coste y el tiempo del proceso de posedición. No obstante, se siguen desarrollando normas generales que se deben aplicar a la posedición de textos, donde tenemos tanto el original como la traducción automática. Por ejemplo en la base de conocimientos de TAUS se sugieren ciertos puntos a considerar durante la tarea de posedición (TAUS 2015) y tener en cuenta ciertos factores previos.

3.4.1 Reglas para la posedición de TAUS

Las recomendaciones que se listan en la base de conocimientos de TAUS (TAUS 2015) se dividen en dos aspectos, que tienen que ver con el momento en que se realice la posedición. Se establecen reglas a tener en cuenta antes de iniciar la posedición y reglas a aplicar durante el proceso mismo de la posedición. Antes de empezar la tarea se deben considerar los siguientes factores:

1. Comprobar si el texto original está bien escrito, editado y preparado para la traducción automática.
2. Averiguar que motor de traducción automática fue utilizado, por ejemplo, si se utilizó un sistema basado en reglas. TAUS también recomienda averiguar si se utilizó una codificación avanzada del diccionario. Si se utilizó un sistema basado en reglas o

3. Posedición

híbrido se debe comprobar si se utilizaron datos de dominio específicos de alta calidad.

3. Determinar qué tipo de posesición se debe realizar.
4. Aclarar con el cliente las expectativas de calidad.
5. Evaluar la calidad de la producción en bruto de la traducción automática, haciendo uso de los diversos indicadores y controles por muestreo.
6. Comprobar si se utilizó algún gestor de terminología.

Una vez se hayan revisado estos aspectos se podrá empezar con la posesición, para la que se proponen las siguientes reglas:

1. La posesición no debe tomar más tiempo que la traducción del mismo texto a partir de cero. La salida de la traducción automática tiene que ser utilizada tanto como sea posible. La posesición de un texto no debe significar la reconversión de otro texto.
2. Para tomar una decisión rápida aplicar "la regla de los 2 segundos", es decir, los poseedores deben ser capaces de decidir en pocos segundos si la salida de traducción automática se puede editar o si se debe retraducir.
3. Palabras sin traducir tienen que ser traducidas a menos que se hayan dejado intencionalmente, como sucede con nombres propios, etc.
4. Segmentos mal traducidos tienen que ser corregidos si las traducciones correspondientes no cumplen el objetivo que supone porque no son comprensibles para el público.
5. Frases inexactas pueden ser borradas, si son difíciles de corregir e irrelevantes en términos del significado de las oraciones originales correspondientes.
6. Factores sintácticos y semánticos son de gran importancia.
7. Puede que no sea necesario corregir errores de estilo en el texto meta; como tal, estas correcciones son objeto de un acuerdo previo, sin embargo, si hay traducciones que pueden ser culturalmente ofensivas, inadecuadas o inaceptables deben ser modificadas.
8. Elementos léxicos recurrentes en el texto meta no deben ser reemplazados con sinónimos, a menos que lo indique un glosario o instrucción específica.

3. Posedición

9. Tienen que ejecutarse medidas de control de calidad general. Traducciones inconsistentes del mismo segmento de texto de origen tienen que ser corregidas. Cambios en las etiquetas tienen que ser corregidos, desajustes numéricos tienen que ser corregidos. Errores terminológicos deben ser corregidos si hay una lista de terminología para el proyecto. Además se recomienda utilizar un corrector ortográfico en el texto meta.
10. Se debe hacer una retroalimentación de los tipos de errores corregidos.

Pero no solo TAUS se ha preocupado por desarrollar reglas y directrices para la tarea de posedición. Existen otro tipo de directrices o reglas de posedición como veremos en los siguientes apartados.

3.4.2 Directrices de posedición para GALE

NIST desarrolló un método de evaluación de la traducción automática usando la distancia de edición como la métrica definida por el programa *Global Language Autónoma Explotación* GALE («Machine Translation Evaluation for GALE» 2006). En este proyecto tuvo varias fases en las cuales se pusieron a prueba traducciones automáticas de texto y datos de voz grabados y además se propusieron las siguientes directrices para la posedición:

1. Hacer que la salida de traducción automática tenga el mismo significado que la traducción humana de referencia. Ni más, ni menos.
2. Hacer que la salida de traducción automática sea tan comprensible como la referencia.
3. Reproducir el significado en el menor número posible de ediciones usando inglés comprensible. Si las palabras, frases o la puntuación en la salida de la traducción automática son totalmente aceptables, deben ser utilizadas; no hacer modificaciones ni sustituirlas por algo nuevo o diferente.
4. La puntuación debe ser comprensible y la oración deben tener puntuación final y el uso de mayúsculas debe ser correcto. No insertar, eliminar o cambiar puntuación; simplemente seguir las reglas tradicionales que determina lo que es “correcto”.

Así como tenemos este tipo de directrices generales, nos encontramos con otras propuestas más concretas que se diferencian según el tipo de posedición que se vaya a realizar.

3. Posedición

3.4.3 Directrices dependiendo del tipo de posesición

En el estudio de Belam (2003) —cuyo propósito es el formular reglas posesición y de esta manera hacer explícita la distinción entre el posesición rápida y mínima. También nos encontramos con reglas para la posesición de TAUS y otras reglas formuladas para la posesición en el ámbito comercial.

3.4.3.1 Reglas para posesición rápida

En este tipo de posesición es más importante el rendimiento que la calidad, por lo tanto el mensaje transmitido debe ser exacto, problemas gramaticales no son importantes, a menos que interfieran con la exactitud, se deben ignorar los problemas estilísticos y no se debe invertir tiempo buscando términos. Tampoco son importantes estándares textuales como cohesión o coherencia. Una guía para la posesición rápida contiene las siguientes reglas:

1. Corregir cualquier palabra que no haya sido traducida.
2. Retraducir cualquier palabra que haya sido traducida de manera errónea y que haga que la frase no tenga sentido.
3. Eliminar cualquier cosa que sea confusa, lo que puede tomar tiempo para corregir y que no agregue, ni deteriore considerablemente el significado.
4. No hacer correcciones de estilo.
5. No cambiar palabras que hayan sido traducidas erróneamente, pero que todavía tienen un significado relacionado.
6. No cambiar el orden de las palabras, si el sentido no se ve afectado.

3.4.3.2 Guías para posesición mínima

El siguiente escalón en términos de calidad del texto poseído sería la posesición mínima. Aquí tiene el poseedor que tener en cuenta las siguientes reglas:

1. Cambiar una elección sorprendente por algo más apropiado y menos llamativo.
2. Cambiar el orden de las palabras para hacer que el texto tenga más fluidez.
3. Cambiar, cuando sea necesario, la estructura de las frases para aumentar la facilidad de comprensión.
4. No gastar mucho tiempo en decidir cuál es la mejor palabra o frase.
5. No intentar hacer que la traducción parezca un texto original.

3. Posedición

3.4.3.3 *Guía para la posedición con calidad “good enough”*

Una traducción con calidad “suficientemente buena” (TAUS 2015) es una traducción comprensible; es decir, se puede entender el contenido principal del mensaje, es precisa, porque comunica el mismo significado que el texto de origen, pero puede ser estilísticamente poco convincente. El texto puede parecer traducido por un ordenador, la sintaxis podría ser algo inusual, la gramática no tiene que ser perfecta, pero el mensaje debe ser preciso:

- Hay que corregir la traducción para que sea semánticamente correcta.
- Hay que asegurarse de que no se haya añadido u omitido accidentalmente ninguna información.
- Se debe editar cualquier contenido ofensivo, inapropiado o culturalmente inaceptable.
- Se debe utilizar, en tanto como sea posible, gran parte de la salida de la traducción automática.
- Se deben aplicar las reglas básicas de ortografía.
- No hay necesidad de aplicar las correcciones que sean solo un carácter estilístico.
- No hay necesidad de reestructurar las sentencias únicamente para mejorar el flujo natural del texto.

3.4.3.4 *Guía para posedición completa*

En una posedición completa el nivel de calidad debe ser tan bueno como si la traducción hubiese sido realizada por un traductor y no un sistema de traducción automática. Este nivel de calidad se define generalmente como comprensible, es decir, un usuario final puede entender perfectamente el contenido del mensaje; la traducción es precisa porque comunica el mismo significado que el texto de origen; además debe ser estilísticamente correcta, aunque el estilo no tiene que ser tan bueno como el que alcanzaría un traductor humano-hablante nativo. La sintaxis es normal, la gramática y la puntuación deben ser correctas. Para que esto se logre se recomienda seguir las siguientes directrices:

- Hay que corregir la traducción para que sea gramatical, sintáctica y semánticamente correcta.
- Hay que asegurarse de que la terminología clave esté traducida correctamente y que los términos no traducidos pertenezcan a la lista que los clientes hayan dado previamente.

3. Posedición

- El poseedor debe asegurarse de que ninguna información se haya añadido u omitido accidentalmente.
- Se debe editar cualquier contenido ofensivo, inapropiado o culturalmente inaceptable.
- Se deben utilizar, en tanto como sea posible, las salidas sin editar de la traducción automática.
- Se deben aplicar las normas básicas de ortografía, puntuación, y la separación de sílabas.
- También hay que asegurarse que el formato sea correcto.

Además de las reglas de posesición que tienen en cuenta el nivel de la posesición, ciertas firmas o empresas de traducción desarrollan sus propias reglas o reglas que se ajusten a un tipo de texto. En el siguiente apartado mostraremos un caso similar.

3.4.3.5 Guías para la posesición en el ámbito comercial

En la tesis doctoral de Tatsumi (2010) nos encontramos con una investigación experimental que desarrolla cierto tipo de reglas de posesición aplicables en un caso particular. Tatsumi (2010, 244) propone reglas que se agrupan según las siguientes preguntas:

A. Qué debe ser arreglado

Tatsumi lista aquí las cuestiones que siempre debe arreglar o corregir el poseedor:

1. Los elementos no traducibles, tales como comandos y nombres de variables nombres, que han sido traducidos.
2. Los términos generales que se hayan traducido inadecuadamente.
3. Las traducciones erróneas (el significado del texto original no se ha transmitido correctamente en la traducción).
4. El orden inapropiado de las palabras, que puede ocasionar que la frase sea imposible o difícil de comprender.
5. Las expresiones comprensibles pero extremadamente antinaturales o inapropiadas.
6. Las posposiciones (en los idiomas en que se emplean) y conjugaciones inapropiadas.

3. Posedición

B. Cuáles elementos deben ser estandarizados

En la tesis de Tatsumi (2010) también se proponen estandarizar ciertos elementos, para hacer el texto más coherente:

1. Los elementos repetitivos, como los procedimientos y pasos.
2. Los estilos de títulos de sección en un manual de usuario y archivos de ayuda.

C. Aquello que no necesita ser corregido

Finalmente Tatsumi también lista los puntos que no deben ser corregidos por el poseedor:

1. La terminación de elementos con viñetas. (Es aceptable tener terminaciones nominales y terminaciones de frases mezcladas entre sí.)
2. Los marcadores de posición, como &ProductNameShort; y &ProductNameLong; ya que serán reemplazados con nombres reales.
3. La puntuación.

El autor da ejemplos que dejaremos de lado porque su combinación lingüística es el inglés y el japonés y no se pueden comprender si no se entiende el idioma japonés.

3.5 Métodos de evaluación del trabajo de posedición

No solo la clasificación de errores, sino también la medición del esfuerzo realizado en el proceso de posedición ya sea en sentido temporal, técnico o cognitivo, nos dan los criterios necesarios para la evaluación de los resultados de la traducción automática. Si bien se han sugerido una serie de estrategias y metodologías sobre la naturaleza de las tareas de posedición humana se necesitaba también de una observación sistemática y de la evaluación objetiva del proceso de mismo de posedición. Krings (2001) realizó un amplio estudio sobre la tarea de posedición y señaló que el esfuerzo en los procesos de posedición no puede ser discutido de manera unidimensional. Krings propone tres aspectos diferentes para medir el esfuerzo en dicho proceso: el aspecto temporal, el cognitivo y el técnico. El esfuerzo temporal es más visible y fácil de medir cuantitativamente, es decir podemos saber cuánto tiempo se tomó el poseedor para realizar una determinada tarea. El esfuerzo cognitivo es la variable determinante que explica el esfuerzo temporal, pero no se puede medir directamente, se tiene que idear

3. Posedición

un método para revelar los procesos mentales de los poseedores. El esfuerzo técnico es el proceso mecánico de la aplicación de las correcciones puede ser descrito como la inserción, eliminación y reordenamiento de palabras. Entre los tres aspectos de la posedición Krings hace un gran énfasis en la investigación del esfuerzo “cognitivo” y por ello lo describiremos a continuación.

3.5.1 *Medición del esfuerzo cognitivo*

Krings emplea el *Think Aloud Protocolo* o *TAP* (Krings 1986) como método para evaluar el esfuerzo cognitivo. Los poseedores deben verbalizar todos los pensamientos que tienen durante el proceso de posedición. Aunque la grabación permite obtener una gran cantidad de información sobre el esfuerzo cognitivo, tiene desventajas, incluyendo 1) el uso de este método en sí crea un ambiente inusual, que pueden distorsionar los datos; 2) el acto de pensar en voz alta puede hacer más lento o alterar el proceso cognitivo normal y 3) no hay garantía de que la declaración verbalizada sea la representación exacta del proceso cognitivo del poseedor (Jakobsen 1998).

También O’Brien (2005) analizó el esfuerzo cognitivo de la posedición combinando dos métodos o herramientas: *Translog* y el análisis de red Choice (CNA). *Translog* es un programa de software que registra todas las pulsaciones de teclas y movimientos del ratón, así como el tiempo empleado. El programa fue desarrollado por Jakobsen (1998), quien partía de la premisa que el comportamiento durante la traducción puede ser analizado cuantitativamente mediante la observación de los procesos técnicos involucrados. Pero también *Translog* tiene una desventaja y es que este método no puede explicar las razones por las que se dan pausas, por ejemplo, si se deben a que el poseedor está pensando o porque le falta habilidad en el manejo del teclado. Es por eso que O’Brien pretende compensar este tipo de desventaja empleando también CNA. CNA (Campbell 2000) es un método para determinar la dificultad de una traducción basada en la diversidad de traducciones realizadas por un número de traductores. Aquí se basa en la idea de que el número de traducciones posibles para un determinado segmento indica la complejidad del proceso cognitivo necesario para realizar una traducción. Con el fin de lograr una medición más directa y detallada del esfuerzo cognitivo en la traducción O’Brien (2007) probó el *eye-tracking*. Ella encontró que el cambio porcentual en la dilatación de la pupila y la velocidad de procesamiento tenían una fuerte correlación.

3. Posedición

Estas métricas pueden ayudar a medir el esfuerzo de posesición humana y a hacer más comprensible la tarea de la posesición, sin embargo no se puede medir unívocamente el esfuerzo cognitivo y ni siquiera podemos hablar de un consenso general en la definición mismo de lo que se considera esfuerzo cognitivo.

Existen diferentes maneras de medir la calidad para la posesición: 1) según el tipo de errores, lo que se puede lograr comparando el texto original con la traducción automática sin corregir; 2) según los cambios realizados: aquí se compara el texto poseído con la traducción automática sin corregir y 3) calculando el esfuerzo, es decir comparando el texto original y la traducción automática sin corregir y estimando cualitativamente el esfuerzo de posesición. Estos métodos son buenos dependiendo del fin que se tenga, es decir son necesarios si se quiere mejorar el desarrollo de los sistemas la medición de errores. Si no solo se pretende mejorar los sistemas, sino también valorar el esfuerzo de posesición, se deben tener en cuenta los cambios realizados (O'Brien 2010).

3.5.2 *Medición del esfuerzo en parámetros de tiempo*

La importancia del factor temporal en términos de medición del esfuerzo de posesición ha sido reconocida desde un comienzo, ya que el tiempo necesario para la posesición influye naturalmente en el coste de la traducción automática (Wagner 1987). El tiempo ha sido empleado en una serie de estudios como medida para determinar el esfuerzo en la posesición, por lo general en forma de velocidad, es decir contabilizando el número de palabras procesadas en un minuto, hora, etc. (véase Krings (2001), O'Brien (2006), Guerberof (2008).

3.5.3 *Medición del esfuerzo técnico*

El esfuerzo técnico puede examinarse desde distintos puntos de vista. O'Brien (2006) mide esfuerzo técnico por un enfoque orientado al proceso utilizando *Translog*, que mantiene un registro de todas las operaciones de teclado y ratón realizados por los poseedores. Groves y Schmidtke (2009) adoptaron un enfoque más orientado al producto mediante el análisis de la diferencia entre los dos textos: uno antes y otro después de la posesición. Compararon la salida de la traducción automática (SMT) y el texto poseído y trazaron el camino mínimo entre los dos. Como resultado del estudio, se enteraron de que las operaciones de posesición más comunes del Inglés al alemán y al francés son inserciones, eliminaciones y alteraciones

3. Posedición

de función las palabras como determinantes y puntuación. Su estudio ha contribuido a identificar patrones de posesión de manera sistemática.

3.6 Herramientas

Existen varias herramientas que pueden ser útiles para la posesión, en el momento algunos entornos o sistemas de traducción como SDL Trados, Wordfast o Déjà Vu X2 integran módulos para la posesión de traducciones semiautomáticas; también algunos sistemas de traducción automática mismos ofrecen la posibilidad de poseer y también nos encontramos con herramientas complementarias desarrolladas especialmente para la tarea de la posesión. En la tabla a continuación listaremos algunas de las herramientas de posesión, ordenadas según las categorías mencionadas anteriormente:

Tipo de herramienta	Nombre
Integrada al entorno de traducción	<ul style="list-style-type: none">• Déjà Vu• MateCAT• MemoQ Translator Pro• OmegaT• SDL Trados Studio• Virtaal
Sistemas traducción automática	<ul style="list-style-type: none">• Apertium Advanced Web Interface• Google Translate• LetsMT• Systran• Tradubi
Herramientas complementarias	<ul style="list-style-type: none">• CheckMate• Kodos• Language Studio Pro• Olifant• PET• Rainbow• RegexBuddy• Symeval• Xbench

Tabla 103: Herramientas para la posesión (Flórez 2012, 221)

Después de la publicación del artículo de Flórez se han desarrollado otras herramientas. Para el análisis de nuestro corpus utilizaremos una de ellas: *translate5* (Mittag 2013). En el marco metodológico la presentaremos más detalladamente.

3. Posedición

3.7 Conclusiones del capítulo

El trabajo de la posedición es indispensable si se emplean sistemas de traducción automática y no es en ningún momento un trabajo de menor importancia que el de la traducción de cero; es otro tipo de actividad que requiere esfuerzo y de las habilidades desarrolladas por el traductor.

Existen varios tipos de ayudas o directrices para los poseedores, que varían de acuerdo al tipo mismo de posedición que se desee realiza. El auge en el desarrollo de la traducción automática también ha contribuido en el desarrollo de herramientas que ayuden al poseedor en la realización de su trabajo.

En el capítulo siguiente expondremos las conclusiones de esta parte de la tesis e introduciremos la justificación de nuestro estudio y la parte metodológica.

4. Conclusiones de la primera parte

Recogemos a continuación las conclusiones del capítulo respetando el orden de los bloques principales que hemos tratado. En el capítulo 2 partimos de los conceptos fundamentales para nuestra tesis. Después de exponer algunas de las definiciones de los conceptos de traducción llegamos a la conclusión que para nuestro trabajo se aplica una definición amplia de lo que se entiende por traducción (véase página 13). Según este tipo de definición la traducción es vista como la creación de un contenido meta que esté de acuerdo con el contenido original según lo convenido en las especificaciones del proyecto de traducción. Esta perspectiva que no está lejos de la funcionalista, pues acepta como correcta una traducción que no sea 100% fluida, siempre y cuando cumpla con los propósitos o la finalidad de los usuarios, considera el proceso de la traducción globalmente, no centrado en el texto y permite incluir en dicho proceso las nuevas tecnologías y la traducción automática. Una definición amplia de la traducción se rige por un concepto de equivalencia pragmático o comunicativo, que se ajuste al *skopo* de la tarea de traducción.

Las definiciones y categorías de problemas de los autores que mencionamos en el capítulo 2 son similares, aunque a veces utilicen otra denominación, y esta similitud deja ver que hay un consenso acerca de la importancia de dichos problemas. Krings y Schmitt denominan los problemas de una manera diferente, pero tienen una clasificación bastante similar. La clasificación de Hurtado también se asemeja a la de Nord. Podríamos afirmar que la definición de problemas latentes según Schmitt enriquece la definición de los problemas culturales o convencionales de Nord, mientras que los problemas que Schmitt define como evidentes los podemos encontrar en parte en la definición de problemas pragmáticos y lingüísticos de dicha autora. Los problemas que Hurtado define como de intencionalidad, es decir aquellos que están relacionados con las dificultades en la asimilación de la información del texto original (intención, intertextualidad, presuposiciones, etc.), son tratados por Nord como problemas pragmáticos.

Pero todos ellos se refieren a los problemas que se pueden encontrar en cualquier tipo de traducción, sin tener en cuenta el contexto concreto del encargo de traducción. La traducción de traducción automática y el proceso de posesición que la acompaña es un tipo de traducción más concreto, generalmente es la traducción de un mismo producto, comparte reali-

4. Conclusiones de la primera parte

dad y por tanto comparte el contexto, por lo tanto hay muchos problemas que —al menos a priori— no se presentan en este tipo de traducción.

Ninguna de las definiciones de estos autores contempla directamente los problemas que pueden presentarse por el uso de motores de traducción automática. Es por esto que complementaremos la lista añadiendo los **problemas específicos de la traducción automática** y haremos una diferenciación dependiendo del motor de traducción automática que se utilice.

Nord	Krings	Schmitt	Hurtado	
Pragmáticos	Comprensión y/o reproducción: <ul style="list-style-type: none"> • del texto origen • de lo comprendido en ambos procesos 	Evidentes	Déficit TP/TT	Lingüísticos
Convencionales			Defectos en el texto de partida	Textuales
Lingüísticos			Déficit de	Extralingüísticos
Específicos del texto	Nivel: <ul style="list-style-type: none"> • palabras • frases • texto 	Latentes	Incongruencias funcionales	Intencionales
			Incongruencias semánticas	Pragmáticos
			Incongruencias semánticas	

Tabla 104: Comparación de las definiciones tradicionales de los problemas de traducción

A continuación definiremos más detalladamente qué tipo de problemas agrupamos en cada categoría que proponemos para nuestra tesis.

1) Problemas pragmáticos

Tanto en la traducción humana como en la traducción automática nos encontramos con problemas pragmáticos que se presentan al traducir un texto que fue producido en una situación original especial y el traductor o el motor de traducción automática debe ajustar la traducción a una situación meta en la que tiene que funcionar de la misma manera que en la situación original. Generalmente el traductor humano debe haber desarrollado la competencia que le permite distinguir si ciertos contenidos y elementos del texto original son específicos de la situación o cultura original y por ello mismo, deben ser adaptados a la situación o cultu-

4. Conclusiones de la primera parte

ra del idioma meta. Se trata de referencias pragmáticas, como datos personales o temporales, deixis de lugar, suposiciones culturales acerca de los conocimientos previos que puede tener el receptor del texto o perspectivas influidas por la cultura. En la traducción automática no se podrá llevar a cabo de manera automática esta adaptación. Los factores externos en la situación original y meta que define Nord (Nord 2001a, 150) que pueden ocasionar problemas de tipo pragmático en la traducción en sentido tradicional son todavía más importantes en la traducción automática. Cierta tipo de referencias temporales deben ser —dado el caso y dependiendo de la situación— adaptadas en la traducción para evitar errores.

Como problemas pragmáticos podemos tratar también los problemas que según Schmitt se dan por déficit en el conocimiento del idioma —tanto del texto original, como el de traducción— así como los que se dan por déficit en el conocimiento especializado sobre el tema del texto a traducir. Krings habla en este sentido de problemas en el proceso de comprensión y en la reproducción del texto a traducir, así como de aquellos que se presentan tanto en el uno como en el otro. Es decir los problemas, tanto de comprensión como de reproducción, que se generan porque la temática del texto original requiere de un conocimiento tecnológico especializado podrían también incluirse entre los problemas pragmáticos.

El incluir los problemas pragmáticos enriquece las clasificaciones actuales de problemas de la traducción automática que se concentran en los aspectos formales o del lenguaje. Si bien este tipo de problemas pueden ser solucionados por un traductor competente, pueden que no sean solucionados en la traducción automática, que funciona más —en el sentido de Nord— como una traducción documental, es decir, como una traducción que trata de reproducir de la mejor manera posible el texto original.

2) *Problemas convencionales o culturales*

Mientras que los problemas pragmáticos se pueden generalizar y la lista que hicimos se puede aplicar para el análisis de éstos en cualquier combinación lingüística, los problemas que tienen que ver con la cultura o las convenciones de una cultura solo se pueden establecer claramente en directa relación con los idiomas y culturas implicadas. Sin embargo podemos ver algunas generalidades, que se presentan en diferentes pares de idiomas. A este tipo de problemas culturales o de convenciones generales nos referiremos en este apartado.

4. Conclusiones de la primera parte

Es importante ver como en una traducción puede cambiar el sentido completamente ya que la cultura y las convenciones también cambian en el transcurso del tiempo.

Es otras palabras problemas culturales surgen por las diversas normas y convenciones que están presentes tanto en la cultura original, como en la de destino (Nord 2001a, 169–190). Existen convenciones que dependen de los diferentes tipos de texto, convenciones generales de estilo, convenciones literarias y formales y otro tipo de convenciones como las que tienen que ver con los sistemas de medidas, que además pueden variar según la cultura. También podemos incluir en los problemas culturales los problemas que Schmitt define como latentes, tanto de tipo funcional, como formal o semántico.

Algunas clasificaciones incluyen los problemas que se pueden dar por la utilización de elementos meta-textuales o supra-elementales como formatos o comillas bajo esta categoría. Estos elementos se utilizan no solo para indicar significados o acentuar una opinión personal, sino también para marcar neologismos que todavía no están reconocidos o para señalar palabras provenientes de otros idiomas. También aquí podemos incluir la utilización de caracteres especiales que no existen en otras lenguas como por ejemplo el signo interrogativo inverso “¿” que utilizamos en castellano al iniciar una frase interrogativa o “¡” al iniciar exclamaciones.

3) Problemas de carácter lingüístico

En cada traducción se enfrentan dos sistemas lingüísticos. Para traducir un texto correctamente el traductor debe que conocer las reglas gramaticales de la lengua meta y ser capaz de ajustar las normas lingüísticas de la lengua origen a las reglas de lengua meta (Nord 2009, 235). Sin embargo debido a las diferencias estructurales de las lenguas pueden presentarse errores en la traducción, y esta problemática se da tanto en la traducción tradicional como en la traducción automática.

Si los problemas o errores se presentan a nivel del significado o léxico, los denominamos como problemas de índole léxico-semántico. Si se presentan porque la estructura de las oraciones es compleja o confusa hablamos de problemas de orden morfológico y sintáctico (Diller y Kornelius 1978, 26–45).

4. Conclusiones de la primera parte

En lo que respecta a los problemas de orden léxico-semántico no podemos negar que uno de los más notables es la existencia de equivalencias de significado que no se pueden resolver simplemente con la ayuda de diccionarios bilingües. El traductor debe desarrollar la competencia que le permita reconocer las relaciones de significación y los posibles sistemas de significación.

Además no podemos negar que los sistemas lingüísticos puedan tener diferentes estructuras semánticas, y que éstas no correspondan 1 a 1 con las categorías que podemos establecer al nivel de contenido. Es por ello que mientras en inglés existen distintas palabras para diferentes tonos del color café o marrón, como lo son *auburn*, *ruddy*, *tan*, *buff* o *dun* la traducción en alemán siempre será una variación del color *Braun*: *rostbraun*, *braunrot*, *Sonnebräune*, *gelbbraun* o *graubraun* (Diller y Kornelius 1978, 29).

Como enunciamos anteriormente, los principales problemas en esta categoría los ocasionan las ambigüedades. Pero entre estas también se pueden hacer algunas distinciones. Hablaremos los siguientes tipos: ambigüedades, homónimos, polisemia, colocaciones o frases idiomáticas y las metáforas y metonimias. En la traducción automática también nos encontramos que este tipo de problemas que se tratan de solucionar —dependiendo del sistema— entrenando o ajustando al motor.

Problemas morfo-sintácticos son aquellos que encontramos en la estructura de los morfemas y las oraciones y que también se dan tanto en la traducción tradicional, como en la traducción automática. Sin embargo los problemas de este orden que se dan en la traducción automática son diferentes a los problemas que se pueden presentar en la traducción tradicional. La ambigüedad estructural y la complejidad gramatical son problemas que suelen generar errores en la traducción automática.

Al abordar la temática de los problemas de la traducción automática hablamos de problemas de orden gramatical además de los problemas de carácter lingüístico. En adelante incluiremos los problemas gramaticales dentro de los problemas lingüísticos de carácter morfo-sintáctico.

4. Conclusiones de la primera parte

4) Problemas que dependen del tipo de texto

Este tipo de problemas es importante para nuestra tesis porque nos dedicaremos al análisis de textos técnicos. La función comunicativa no es solo el rasgo constitutivo básico de los textos, sino también el marco según el que se producen del tipo de texto, es decir la función comunicativa del texto determina su tipología. “Todos los tipos de textos que poseen rasgos muy similares en una lengua pertenecen al mismo género, y éste funciona como una especie de prototipo textual o de abstracción del formato que debe tener un texto” (Gamero 2001, 49). Es decir, en estos tipos o géneros se encuentran repetidamente ciertas características, lo que se constituye en una especie de señal que facilita a los participantes en el proceso comunicativo reconocer las funciones comunicativas previstas y por lo tanto entender el texto de la manera adecuada.

Las convenciones que caracterizan un tipo de texto son los rasgos textuales utilizados por los hablantes de un idioma, para comunicar de una manera un contenido, que podría ser transmitido correctamente de otra manera (Gamero 2001, 51). Por ejemplo en un manual de instrucciones no se suele utilizar el presente indicativo en segunda persona singular, a pesar de que su uso sería gramaticalmente correcto. Sin embargo dichos rasgos pueden ser diferentes dependiendo del idioma. Por ello aquí puede haber problemas en la traducción automática, si se traduce el original de acuerdo con las convenciones tipológicas de éste y no se adaptan al lenguaje meta.

Como en nuestra tesis nos centraremos en la traducción de textos técnicos no nos referiremos por a algunos de los problemas que se dan en la traducción de otro tipo de textos como la traducción de redundancias u otros recursos estilísticos como los enlaces asindéticos. Problemas que dependen del texto y que son relevantes para nuestra tesis son aquellos que tienen que ver con el léxico, la terminología y las convenciones de género empleadas.

5) Problemas en el texto original

De mayor importancia para la traducción automática que para la traducción tradicional son los problemas que pueden presentarse por errores en el original o defectos en el texto original por ejemplo cuando el texto no está completo, es ambiguo o hay incoherencias. Nos referimos al tipo de problemas que surgen por lo que Schmitt denomina defectos en el origi-

4. Conclusiones de la primera parte

nal, como los defectos en los aspectos formales del texto (p. ej. si un párrafo se repite, si falta una parte o si una referencia es errada) que no pueden ser detectados por los motores de traducción automática. También nos encontramos aquí los problemas de traducción que se pueden dar por errores en las cifras o unidades de medida y los problemas que pueden ocasionar errores de ortografía o de imprenta.

Schmitt también enumera ciertas discrepancias en el texto original que pueden ocasionar problemas de traducción. Este tipo de discrepancias son también problemáticas en la traducción automática, donde es aún más difícil identificarlas. Nos referimos a aquellas que se pueden dar cuando una imagen se le asigna un texto que no le corresponde o los términos empleados para su descripción no se ajustan a la terminología empleada en la parte textual (Schmitt 1999, 83–85).

También es problemático si se dan errores o defectos en la formulación del texto original que hacen difícil su comprensión. Este tipo de problemas puede presentarse en la traducción de textos técnicos en los que no es fácil formular instrucciones o explicaciones, respetando las convenciones del tipo de texto, de manera comprensible para el receptor sin omitir o añadir información.

6) *Problemas específicos de la traducción automática*

Muchos de los problemas que se enumeran en las clasificaciones tradicionales se dan también en la traducción automática pero el grado de importancia puede ser diferente.

Problemas típicos de la traducción automática son, por ejemplo, las omisión y adición de palabras e incluso partes de segmentos, errores en el orden gramatical y como lo mencionamos en el capítulo “Problemas lingüísticos” a partir de la página 53 aquellos problemas que se presentan cuando no se resuelven correctamente las ambigüedades tanto léxicas como estructurales. Además también se presentan problemas de orden pragmático, es decir aquellos que se dan cuando no se cumple la función comunicativa del original (véase página 31).

Para la evaluación de los textos de nuestro corpus la clasificación de los problemas y de los errores que se dan cuando estos problemas no se resuelven o solo se solucionan parcialmente será una guía. En el análisis podemos comprobar si en un caso real encontramos

4. Conclusiones de la primera parte

estos problemas o si aparecen otros. Del mismo modo si los problemas que en el marco teórico son descritos como importantes, en un caso real también tienen la misma importancia y podremos valorar cuantitativamente su importancia.

Si solo se quiere entender de qué se trata el texto, sin importar si la traducción es correcta gramatical o culturalmente, la traducción que cumpla con este objetivo sería correcta. En este sentido la traducción automática podría calificarse como correcta si no presenta errores funcionales, es decir en tanto podamos entender de qué se trata —por ejemplo en el sentido del *gisting*— y es por ello que definición funcional o comunicativa de lo que es una traducción correcta podría ser demasiado tolerante para la evaluación de los problemas de la traducción automática.

Partimos de la premisa de Hurtado en la que afirma, que cualquier tipificación de errores persigue solo la búsqueda de categorías que proporcionen un metalenguaje que ayude en el análisis, pero que no dan cuenta por sí solas de la gravedad o incidencia del error ni tienen un valor explicativo (Hurtado 2011, 307). También recalcamos el hecho de que solo tendremos en cuenta los errores de la traducción escrita. No consideraremos, ni tendremos en cuenta para esta tesis las relaciones entre los errores y la competencia traductora, ni buscaremos en ésta última la causa de ellos. Las tipologías de errores que describimos en el marco teórico nos ayudarán en la parte metodológica en el análisis los textos del corpus, y allí podremos comprobar si la lista se ajusta o si encontramos otro tipo de problemas que no estaban incluidos en nuestra clasificación inicial.

Para ello tendremos en cuenta la clasificación de inadecuaciones de Hurtado, aunque no hablaremos de inadecuaciones sino de errores. En este sentido hablaremos de errores cuando:

1. Estos afecten la comprensión del texto original. Es decir cuando se de un sentido falso, cuando la traducción no tenga sentido, no tenga el mismo sentido; cuando se den adiciones u omisiones; cuando no se solucionen correctamente las referencias extralingüísticas y se de una inadecuación de la variación lingüística (tono, dialectos, idiolecto, etc.).

4. Conclusiones de la primera parte

2. También hablaremos de errores cuando estos afecten la expresión en la lengua meta: es decir aquí nos encontraremos con errores de ortografía, puntuación, gramática, léxico, aspectos textuales (coherencia, progresión temática, elementos de referencia, conectores) y errores de redacción (formulación defectuosa o poco clara, falta de riqueza expresiva, pleonasmos, etc.)
3. Errores pragmáticos que no son acordes con la finalidad de la traducción (en relación con el tipo de encargo o destinatario), el método elegido, el género textual y sus convenciones, etc. (Hurtado 2011, 305–306).

A estas tres categorías agregaríamos, teniendo en cuenta la clasificación de errores de Schmitt, los errores que se producen por defectos en el texto de partida. Estos errores son:

4. Errores —o defectos siguiendo la denominación de Schmitt— formales, como por ejemplo repeticiones u omisión de partes o palabras en el texto original. Incluiremos también en esta categoría los errores ortotipográficos, a pesar de que el autor los considere diferentes.
5. Errores en las cifras o unidades de medida.

No tendremos en cuenta los niveles en que se presentan los errores, es decir si se dan a nivel de las oraciones o párrafos, porque en el corpus solo tendremos acceso a traducciones de segmentos.

Como nos referiremos en nuestra tesis particularmente a la traducción automática; por ello también tendremos en cuenta para la clasificación de los errores aquellos errores que ocurren al traducir utilizando motores de traducción automática. Para evaluar este tipo de errores la distinción de las categorías que encontramos en la tipología de errores de Loffler-Laurian es interesante, ya que en esta clasificación se encuentran incluidas las clasificaciones de los otros autores mencionados. Sin embargo no descartamos la posibilidad que en el la parte metodológica de nuestra tesis encontremos otros tipos de errores que no se hayan incluido en ninguno de los esquemas mencionados.

Para el análisis de nuestro corpus no son relevantes algunas de las categorías que tienen en cuenta ciertas tipologías de errores, como aquellas que tienen que ver con la autoedi-

4. Conclusiones de la primera parte

ción o la interfaz de usuario (véase página138), no porque no los consideramos importantes sino porque en los textos del corpus no tenemos acceso a los textos autoeditados, ni a las interfaces de usuario de los mismos.

Consideramos que todos los errores que se pueden dar en diferentes niveles, es decir en las palabras, las relaciones o la estructura de la información pueden ser funcionales o absolutas. El que sea de una u otra categoría solo se puede determinar en cada caso en particular.

La distinción que hace Loffler-Laurin (véase página 130) de diferentes niveles en que se presentan los errores es interesante además porque no pretende dar una clasificación valorativa de los errores, ya que un mismo error puede tener mayor o menor implicación en dos segmentos diferentes, y en una oración pasar desapercibido, mientras que en otras puede cambiar completamente el sentido.

4. Conclusiones de la primera parte

PARTE II: MARCO METODOLÓGICO

El marco metodológico está dividido en cuatro capítulos. En primer lugar, el capítulo 5 en el que presentaremos nuestro objetivo; en el capítulo 6 describiremos el corpus, su estructura original, la manera en que reorganizamos los datos y las herramientas adicionales que utilizamos. En el capítulo 7 encontramos los resultados del análisis preliminar de todos los datos del corpus y una descripción global de los errores encontrados según al motor y según categorías gramaticales. Al final del capítulo hicimos una revisión de los datos según las categorías de MQM.

En el capítulo 8 encontramos la clasificación que proponemos para la sistematización de la posesión en nuestra combinación lingüística. Allí describimos la manera en que procesamos los datos, las definiciones de macrocategorías, categorías y valores que empleamos para describir las ediciones que realizaron los poseedores.

5. Presentación de los objetivos

En este capítulo introductorio del marco metodológico nos proponemos, una vez fijado el marco teórico, enumerar los objetivos de esta tesis y también exponer la estructura de este marco.

5.1 *Objetivos*

El principal objetivo de esta tesis consiste en proporcionar una visión pormenorizada de la posesición de textos técnicos del alemán al español a partir del análisis de un corpus de segmentos originales y traducidos automáticamente. El cumplimiento de este objetivo permitirá identificar criterios que favorezcan la creación de estrategias para la formación de poseedores teniendo en cuenta tanto la combinación lingüística como las circunstancias habituales de un encargo de posesición. De este modo, los resultados de esta tesis se podrán convertir en un punto de partida para la formación académica de poseedores.

Para cumplir con este objetivo debemos plantearnos una serie de objetivos previos o parciales, que serían:

1. Describir el corpus de segmentos originales y sus traducciones con cuatro motores diferentes.
2. Analizar los resultados de cada motor.
3. Establecer una categorización de errores de acuerdo con las ediciones introducidas durante la fase de posesición, así como con su segmento origen.
4. Analizar los resultados teniendo en cuenta las categorías de errores establecidas y el motor de traducción automática que produjo el segmento elegido por el poseedor.

Durante la posesición se editan los segmentos producidos con traducción automática para garantizar una traducción correcta del original. En el marco teórico vimos que ya hay varias tipologías y métodos de clasificación de errores de traducción (independientemente de si se trata de traducción humana, asistida por ordenador o automática), que abordan la temática desde el punto de vista del texto de llegada; nuestro objetivo es abordar los textos y los errores de la traducción automática teniendo en cuenta también los textos de partida y viendo cómo se relacionan (Objetivo 1). En este sentido, nuestro enfoque difiere de otros sistemas de

5. Presentación de los objetivos

caracterización de errores o de análisis de la calidad (que tienen en cuenta únicamente el texto de llegada) porque nuestro interés no es mejorar los resultados de los sistemas de traducción automática sino hacer un aporte en la formación de traductores y poseedores.

Queremos ver si podemos realizar una clasificación de errores que ayuden al poseedor en nuestra combinación lingüística (Objetivo 2); pretendemos analizar si entre los errores y los motores empleados existe algún tipo de relación que pueda aportar algún tipo de ayuda en la formación de traductores o poseedores con la información obtenida. Como indicamos en la introducción de nuestra tesis el objetivo es profundizar en la descripción los errores en que se dan en la traducción automática en textos de carácter técnico para facilitar el establecimiento de estrategias y técnicas de posesición.

Para ello describiremos en el capítulo 6 en primer lugar el corpus, su estructura y en segundo lugar la manera cómo procesamos los datos. En el capítulo 7 presentaremos el primer análisis de los errores encontrados teniendo en cuenta todas las propuestas del corpus. En este capítulo detallaremos las diferencias entre las traducciones obtenidas con cada uno de los motores de traducción automática. En el capítulo 8 presentaremos el segundo análisis en el que solo tuvimos en cuenta los segmentos seleccionados para la posesición. Allí también describiremos en primer momento el subcorpus obtenido y la manera en que procesamos los datos. Aquí analizamos los errores que se han editado en la posesición y no el proceso mismo de la posesición. En este capítulo también presentaremos nuestra propuesta de clasificación de errores en seis macrocategorías (Objetivo 3).

Al final de esta parte esperamos ver qué resultados nos ofrecen los datos por macrocategorías y vamos a ver qué tipos de errores encontramos y si es posible clasificarlos teniendo en cuenta las clasificaciones mencionadas en el marco teórico o si es necesario realizar otro tipo de clasificación (Objetivo 4). También analizaremos si los errores encontrados se pueden vincular con algún tipo de motor concreto: si su distribución es aleatoria o, a la inversa, si hay motores concretos que nos van a dar algunos tipos errores especiales.

5.2 *Aplicación de sistemas de medición de la calidad como instrumento de investigación: el caso de MQM*

Para resaltar la diferencia de nuestra propuesta con las propuestas existentes hemos incluido en este capítulo la descripción de un sistema de medición de calidad. En el capítulo 3.7.2 del marco teórico introdujimos MQM como uno de los métodos de evaluación de la traducción automática con intervención humana. MQM es una métrica (Multidimensional Quality Metrics) que proporciona un marco para describir y definir indicadores de calidad que se pueden utilizar para evaluar la calidad de los textos traducidos y para identificar problemas específicos en esos textos. MQM se ajusta al enfoque funcionalista que define una traducción como buena en tanto cumpla con su propósito comunicativo y es un marco teórico para definir métricas de acuerdo a las especificaciones concretas de un proyecto. Es por ello que MQM no se puede describir como una métrica estática y única que proporciona criterios universales que pueden utilizarse para todo tipo de traducciones y pueden emplearse para la evaluación de cualquier proyecto de traducción. Más bien MQM ofrece —como dijimos anteriormente— un enfoque funcional de la calidad que busca ver si se cumplen o no las especificaciones concretas de una traducción. Se puede aplicar a textos monolingües o bilingües. Las categorías empleadas en MQM han sido definidas a partir del análisis de otros tipos de categorías como SAE J245, TAUS DQF y LISA.

Con el marco multidimensional que nos proporciona MQM se pueden describir y definir indicadores de calidad para evaluar textos traducidos y para identificar problemas específicos en dichos textos y además MQM permite la utilización de dicho marco sistemático para describir y crear métricas de calidad para un determinado proyecto o contexto. Su aplicación se orienta fundamentalmente a la evaluación del resultado de un proceso de traducción: el texto final. Este es el principal motivo por el que no podría satisfacer por completo el objetivo planteado en esta tesis. Aun así, nos hemos propuesto profundizar a continuación en su aplicabilidad y describir el corpus también mediante este marco de evaluación (apartado 7.7).

Veamos de qué elementos se compone dicho marco:

1. En primer lugar MQM nos proporciona un vocabulario para categorizar problemas de calidad y nos da una definición de tipos de errores (*issue types*).

5. Presentación de los objetivos

2. También tiene un mecanismo de puntuación con los que se pueden definir los niveles de calidad alcanzados, que se basa en el conteo de errores o en las anotaciones sobre errores en particular (*scoring*).
3. Nos da una estructura (*markup*) para describir la métrica de MQM o crear una métrica compatible con las definiciones y con la jerarquía de la tipología de errores. Esta estructura incluye:
 - Un mecanismo formal en XML para declarar o describir las métricas de calidad. Es decir, un vocabulario que permite a los ejecutores y a los usuarios describir las métricas o emular las métricas en formato de MQM (*metrics description*).
 - Un conjunto de atributos en el `<mqm:namespace>` que se pueden utilizar con XML o HTML 5 para insertar los datos de MQM en estos formatos de archivo (*MQM inline attributes*). Aquí hablamos de elementos como `issueType` o `issueSeverity` que a su vez se puede especificar en `minor`, `major` o `critical`.
 - Un conjunto de elementos en el `<mqm:namespace>` que se puede utilizar para insertar datos en archivos XML cuando los elementos existentes no cumplen los requisitos (*MQM inline elements*).
4. Un conjunto de directrices para la selección de errores basado en la norma *ASTM F2575: 2014 specification*.
5. Un conjunto de las aplicaciones informativas sobre las modificaciones actuales en las métricas de MQM que puede resultar útil para los evaluadores, para seguir los cambios realizados en cada versión.

Para nuestra tesis no utilizaremos todos los elementos del marco; para nuestro trabajo es en primer lugar importante la definición de errores y los mecanismos de evaluación. A continuación entraremos en detalle en estos puntos.

5.2.1 Definición de calidad y errores en MQM

Antes de describir detalladamente la clasificación de errores de MQM debemos indicar cómo se define en este contexto una traducción de calidad y cuál es la definición de error con que trabaja MQM. Una definición universalmente válida de calidad no existe, ya que la

5. Presentación de los objetivos

calidad es vista como algo relativo, que depende de varios factores como el propósito, la audiencia etc. Es por ello por ejemplo que en el proyecto QTLauch Pad (2014), que ha desarrollado MQM, se habla de una definición multidimensional de calidad. Según esta perspectiva una traducción de calidad es aquella que demuestra la precisión y fluidez requerida para la audiencia y el propósito específico y que cumple con todas las especificaciones negociadas anteriormente, teniendo en cuenta las necesidades del usuario final.

Un error representa cualquier problema que se pueda encontrar con el texto traducido que, o bien no corresponde con la fuente o se considera incorrecto en el idioma de destino. MQM define *issue types*, solo un humano puede determinar si estos se presentan o no. Con base en un análisis detallado de los indicadores de calidad de la traducción, MQM proporciona una tipología flexible de tipos de error que puede aplicarse a tareas de evaluación analítica o integrales de evaluación de calidad de la traducción. La lista de errores de MQM se divide en dos categorías principales: precisión y fluidez; cada una de estas categorías tienen subcategorías más detalladas. Sin embargo MQM define otras categorías que pueden ser tenidas o no en cuenta dependiendo del proyecto y de las especificaciones de éste y las organiza jerárquicamente⁷.

5.2.2 Tipos de errores

Como enunciamos anteriormente las categorías de MQM no son estáticas y pueden ser incluso modificadas para su utilización en un proyecto de evaluación concreto. Si bien se recomienda a los usuarios limitar los tipos de errores a los predefinidos ya por MQM, se permite —en caso necesario— agregar tipos de errores adicionales para cumplir con requisitos especiales de un proyecto. Las categorías que se añadan tienen que incluir la siguiente información:

- *Name*: un nombre que defina el problema
- *ID*: un nombre sin prefijo que sirva como identificador del error en XML. Debe empezar con x- para indicar que es una categoría que ha definido el usuario

⁷ También debemos anotar que la clasificación y jerarquización de errores de MQM puede variar dependiendo de la versión. La tipología que utilizaremos es la presentada en la versión de diciembre del 2015.

5. Presentación de los objetivos

- *Parent*: El valor del ID del tipo de error del que proviene. Este puede ser un tipo de error de MQM o uno definido por el usuario
- *Definition*: la definición del tipo de error.

5.2.3 Scoring

El modelo de puntuación MQM se aplica solo al conteo de errores de las implementaciones de MQM. En la actualidad esta especificación no define un modelo de puntuación por defecto para los sistemas holísticos, que son menos detallados en la naturaleza que la métrica de error de conteo. Las futuras versiones, sin embargo, pueden definir un modelo predeterminado de sistemas holísticos.

El modelo de calificación proporciona un método para calcular un único nivel de calidad como un valor porcentual. Dichas puntuaciones se utilizan con frecuencia para las pruebas de aceptación en los procesos de aseguramiento de la calidad de la traducción. Además, se pueden obtener otro tipo de puntuaciones derivadas dependiendo de los diversos aspectos del tanto del objetivo como del texto de original; también se pueden implementar otros tipos de calificación. Se plantean niveles de gravedad por defecto que tienen en cuenta el *Weight* y la *Severity*. Todos los temas tienen un peso predeterminado de 1,0. Este peso puede ser actualizado en función de cada tema para reflejar los requisitos específicos. Los niveles de gravedad por defecto se definen como sigue:

- ***None 0***: items que no tienen que ser marcados y no se deben contar: Este tipo puede usarse como una marcación que no imponga sanción.
- ***Minor 1***: para marcar pequeños problemas que no afectan a la usabilidad o la comprensibilidad de los contenidos
- ***Major 10***: Las principales cuestiones son cuestiones que la usabilidad de impacto o la comprensibilidad de los contenidos, pero que no inutilizarla
- ***Critical 100***: Los temas críticos son cuestiones que hacen que el contenido no apto para su uso. Por ejemplo, un error gramatical particularmente malo que cambia el significado del texto se considera crítico. Si el error impide que el lector/usuario de usar el contenido de lo previsto o si se presenta información incorrecta que podría ocasionar daños al usuario que deberá ser categorizado como crítico.

5. Presentación de los objetivos

MQM brinda la posibilidad no solo de categorizar el error, sino también de establecer su gravedad en tres categorías: mayor, menor y crítico. Sin embargo en la práctica la determinación de un error como crítico no es recomendada.

5.2.4 Utilización de MQM

La conformidad de una métrica de evaluación de la calidad de traducción con MQM está determinada por los siguientes criterios:

- La métrica debe utilizar el vocabulario de tipos de error MQM.
- Cualquier tipo de error utilizado por la métrica que no esté contenido en MQM debe definirse según extensiones del usuario MQM (*MQM user extensions*) y tener un lugar apropiado en la jerarquía de MQM.
- Tipos de errores definidos por los usuarios no pueden reemplazar tipos de errores ya existentes en la definición de MQM que tengan el mismo significado. Por ejemplo, una métrica que defina como tipo de error “Not translated” tiene el mismo significado que la categoría de MQM “untranslated” y por lo tanto el tipo de error “Not translated” no es conforme con MQM.
- La métrica puede usar el `<custom displayName>` para asignar un nombre no predeterminado a un tipo de error. Por ejemplo, una métrica podría declarar por ejemplo un elemento como *Not translated*, por ejemplo como el nombre a visualizar para el tipo de error *Untranslated*. En los casos en que se usen diferentes nombres se debe utilizar el ID de MQM utilizar para evitar confusiones.
- La métrica debe utilizar un tipo de marcado MQM declarado.
- La métrica que pretenda ser conforme a MQM debería definir los errores teniendo en cuenta la jerarquía MQM, pero no está obligada a hacerlo.

5.2.5 Cómo crear una métrica con MQM

Cuando se utiliza MQM para evaluar la calidad de la traducción en realidad se está usando una métrica que es compatible con MQM. Para crear una métrica que sea compatible con MQM se deben determinar en primer lugar los tipos de errores a chequear y el nivel de granularidad de la evaluación.

5. Presentación de los objetivos

La métrica compatible con MQM más general es la que identifica tan solo dos categorías de errores: precisión y fluidez; incluso si solo se controla el texto traducido basta con el criterio de fluidez. Los criterios de evaluación pueden variar de acuerdo a las especificaciones del proyecto a evaluar. En general, una métrica compatible con MQM diseñada para una tarea de investigación debe tener las siguientes cualidades:

1. Debe ser lo suficientemente granulada para abordar las cuestiones de investigación pertinentes.
2. La métrica no debe contener categorías extrañas o pedir a los correctores marcar temas irrelevantes en la investigación.
3. La métrica (los tipos de errores utilizados) debe ser lo suficientemente compacta como para permanecer en la memoria del anotador.
4. El anotador tiene que tener ayuda para poder solucionar casos ambiguos.

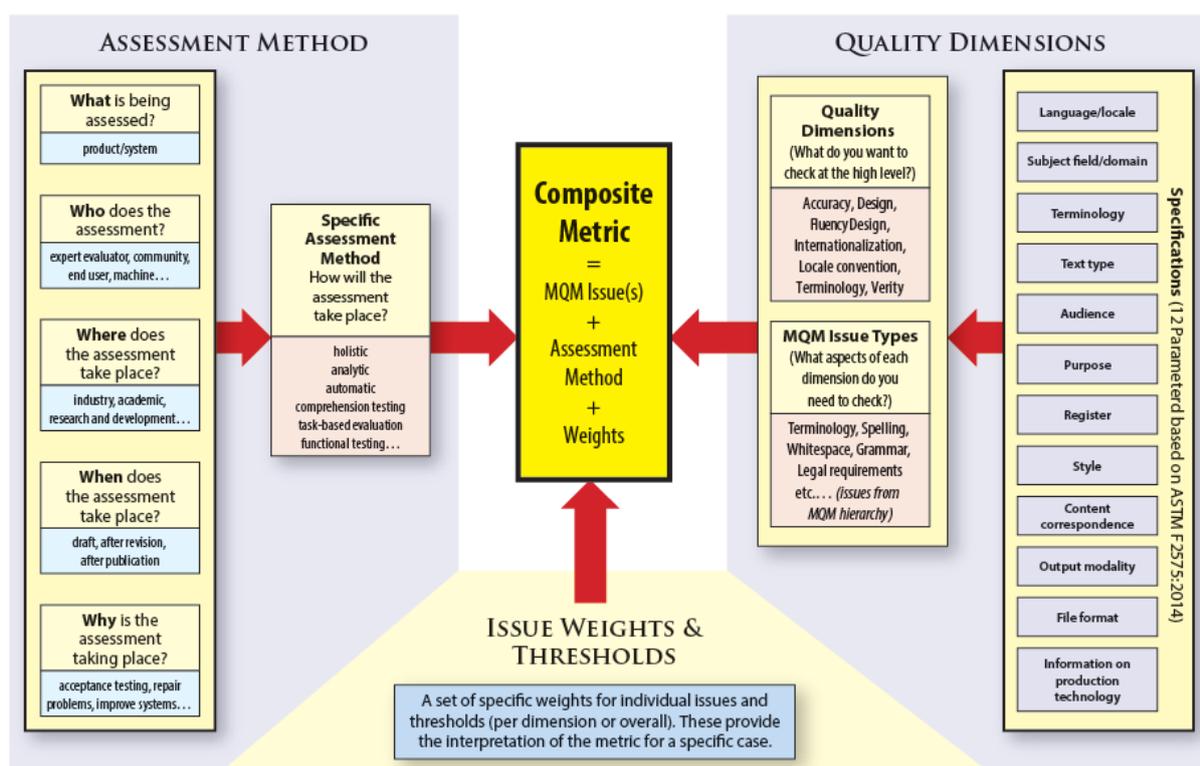


Ilustración 26: Esquema de relación de MQM («MQM (Multidimensional Quality Metrics)» 2014)

5. Presentación de los objetivos

Lo primero es determinar qué tipo de métrica se va a utilizar, si se trata de una métrica analítica, holística, basada en el texto o funcional, etc. Ello se puede definir teniendo en cuenta las siguientes cuestiones:

- Qué se está evaluando: un texto traducido, el proceso de traducción, un software.
- Quién hace la evaluación: traductores expertos, lingüistas, usuarios, etc...
- ¿Dónde se llevará a cabo la evaluación: en un ambiente académico o de investigación?
- Cuándo se llevará a cabo la evaluación: en qué etapa se encuentra el proceso, se ha revisado la traducción o no, etc.
- Por qué se realiza la evaluación: para determinar si la traducción puede ser publicada, para mejorar problemas, para evaluar un sistema, etc.
- Con base a las respuestas a estas preguntas se puede determinar qué tipo de análisis es mejor:

Analítico: enfoques analíticos implican la identificación de problemas específicos en el texto. Estos enfoques tienen en cuenta los detalles y son especialmente adecuados para la detección y reparación de errores individuales. La evaluación no debe ser realizada por usuarios finales, ya que necesitan de un conocimiento o una formación.

Holístico: estos enfoques consideran el texto en su conjunto. En lugar de identificar errores específicos en el texto evalúan la calidad general del texto con respecto a los criterios específicos. Hay dos subtipos de métodos de evaluación holística: simples/mono dimensionales o complejos/multidimensionales. En el primer enfoque se evalúa el texto de acuerdo a un solo criterio. Este tipo de enfoque es apto para la evaluación por parte de usuarios finales. En el otro tipo se evalúa el texto dependiendo de varios criterios que se han definido de ante mano. Dicha evaluación puede ser apropiada si se prefiere que el texto sea evaluado por varios evaluadores, con una formación diferente. Ente más precisos y mayor sea el número de criterios de evaluación, también tendrá que ser mayor la capacitación que se le debe dar a los evaluadores.

Task-based testing. Las pruebas basadas en tareas evalúan qué tan bien la traducción sirve para ayudar a los usuarios a realizar una tarea. Es particularmente útil para determinar

5. Presentación de los objetivos

problemas con la categoría *Verity*. Las evaluaciones basadas en la tarea es una disciplina por derecho propio y no está incluida en MQM, aunque aspectos de la prueba basada en tareas pueden estar relacionados con los tipos de errores de MQM.

Functional testing. Las pruebas funcionales evalúan si una traducción funciona como se esperaba. Es más adecuado para evaluar proyectos de traducción u otros productos que combinan traducción con otras funciones. Las pruebas funcionales van más allá del alcance de MQM, aunque, al igual que con las pruebas basadas en tareas, pueden estar relacionadas con algunos de los tipos de errores de MQM.

Además de seleccionar un método de evaluación basado en las respuestas a las preguntas de la izquierda del diagrama, los usuarios también tienen que definir las especificaciones (es decir, los valores de los parámetros) para la traducción que deben evaluarse. Para ello se pueden tener en cuenta los parámetros que establece MQM (12) y Sobre la base de las especificaciones, los usuarios deciden qué se evaluarán dimensiones del texto. Dimensiones definidas en MQM son los tipos de errores:

Las métricas analíticas se crean al hacer una selección de temas relevantes de la lista de MQM tipos de emisión. El siguiente procedimiento puede utilizarse para crear una métrica:

Completar un conjunto completo de especificaciones del proyecto, incluyendo los 12 parámetros de MQM. Asegurarse de que todas las partes están de acuerdo sobre los valores de los parámetros. En la tabla a continuación listamos los parámetros con una breve descripción. MQM utiliza una selección de 11 de los 21 parámetros definidos en ASTM f2575 y adiciona uno (*Output modality*):

Parámetro	Descripción
1. <i>Language/locale</i>	Idioma en que el texto se ha traducido y si hay una variante regional, por ejemplo alemán de Suiza.
2. <i>Subject field/domain</i>	Campo(s) o dominio(s) del original, p.ej es un texto legal del siglo XVII
3. <i>Terminology</i>	Lista de términos o referencia a términos que se utilizarán.
4. <i>Text type</i>	Tipo de texto del original: manual de usuario, novela...

5. Presentación de los objetivos

Parámetro	Descripción
5. <i>Audience</i>	A quién está destinado el proyecto: usuarios de una tableta, analistas financieros, etc.
6. <i>Purpose</i>	El propósito del proyecto: está destinado a transmitir información, a persuadir el público.
7. <i>Register</i>	Descripción del registro lingüístico a utilizar en el idioma de destino: conversación informal, carta legal, etc.
8. <i>Style</i>	Especifica información sobre el estilo y su importancia.
9. <i>Content correspondence</i>	Especifica cómo se va a traducir el contenido
10. <i>Output modality</i>	Información sobre la forma en que se va a presentar el texto traducido (online, display de un móvil.
11. <i>File format</i>	El formato o formatos de archivo en el que el contenido traducido debe ser entregado.
12. <i>Production technology</i>	Cualquier tecnología o software para ser utilizados en el proceso de traducción, p.ej., si se usó una MT o una base de datos de terminología

Tabla 105: Parámetros de MQM

- Para fijar el valor de cada parámetro, hay que tener en cuenta qué características del texto se deben verificar para determinar que el texto cumpla con las especificaciones y tenga en cuenta estos tipos de error. Tenga en cuenta que "no importa" es un valor aceptable para muchos parámetros y si se elige este valor, el parámetro se puede omitir. (Por ejemplo, si el estilo se juzga como irrelevante, a continuación, este parámetro se pasa por alto en la evaluación).
- Después de decidir qué debe ser verificado, determinar qué tipos de errores pueden ser utilizados para la evaluación.
- Priorizar la lista de errores basándose en la importancia de cada parámetro y luego hacer una selección de tipos de errores basándose en esta lista y en las prioridades establecidas. Seis a siete tipos de errores son más que suficientes para la tarea de evaluación.
- Determinar si se va a evaluar la gravedad de los errores y a asignarles un 'peso'. Si la métrica resultante debe ser implementado en una herramienta compatible con MQM, debe ser declarada como lo describen sus autores (Lommel et al. 2016).

5. Presentación de los objetivos

Los métodos de evaluación holísticos son más flexibles en algunos aspectos que las métricas conteo de errores. Están diseñados para proporcionar una evaluación del texto traducido en su conjunto en lugar de una contabilidad detallada de todos los errores. Dado que la evaluación analítica puede llevar mucho tiempo y no es necesaria en todos los casos (por ejemplo, cuando la pregunta es si un texto se debe aceptar o no), los métodos holísticos pueden ser más apropiadas solo en algunos casos. La mayoría de los tipos de errores de MQM se puede utilizar fácilmente como cualquiera de los tipos de análisis. Por ejemplo, la puntuación de MQM según errores puede ser utilizada por pedir evaluadores usando una herramienta integral si el texto está marcado correctamente. En este contexto, algunos problemas serán más útiles que otros.

La perspectiva desarrollada para la métrica de MQM es sin duda muy interesante para futuros estudios, desde nuestra perspectiva y teniendo en cuenta nuestros objetivos es una aproximación interesante. En nuestra categorización de errores, como veremos en el capítulo 8, no daremos puntuación ni hablaremos de niveles de gravedad. Tampoco priorizaremos la lista de errores encontrados, ni utilizaremos patrones predefinido, pero si utilizaremos en parte las categorías en que se definen los errores según MQM, tres de cuatro criterios que también se utilizan en la métrica de MQM: Nombre de la categoría, identificado y definición.

Después de esta introducción a nuestros objetivos generales seguiremos en el capítulo siguiente con la descripción del corpus de nuestra tesis.

6. Descripción del corpus y de las herramientas de análisis

Como punto de partida tenemos básicamente dos aspectos fundamentales: el análisis descriptivo del corpus y la evaluación de errores con ayuda de la tipología de errores de MQM. En primer lugar describiremos nuestro corpus y la manera en que lo organizamos para el análisis. En segundo lugar haremos una descripción de los tipos de errores de MQM y de las herramientas que tuvimos a disposición para su utilización. No vamos a darle peso a los errores porque para poder clasificar un error como más grave o no también necesitaríamos más información acerca del contexto y nuestro enfoque está centrado en analizar lo que se ha corregido en la posesición y no en si el error fue de mayor o menor peso.

6.1 Descripción del corpus

El corpus con que trabajamos en nuestra tesis es un extracto del corpus del proyecto TaraXÜ («DFKI LT - taraXÜ» 2015). Este proyecto contó con la participación de investigadores del DFKI (*German Research Center for Artificial Intelligence*), así como con la participación de investigadores y profesionales de la traducción de proveedores de servicios de traducción como Euroscript Deutschland GmbH, Acrolinx GmbH y yocoy GmbH. El proyecto se realizó con la financiación del TSB (*Technologie Stiftung Berlín*) y fue cofinanciado por el Fondo Europeo de desarrollo regional de la Unión Europea EFRE (*European Fund for Regional Development*).

El proyecto, destinado primordialmente a desarrollar un nuevo sistema de evaluación y selección del mejor motor traducción automática para un cliente particular, se inició en abril de 2010 y fue culminado en septiembre de 2013. Durante el proyecto los investigadores realizaron un estudio comparativo y paralelo de los principales modelos de traducción automática actuales: aquellos basados en patrones, en reglas o en estadísticas y utilizaron también una memoria de traducción.

Los análisis realizados durante el proyecto también permitieron identificar problemas en los sistemas de traducción automática que, según los participantes, podrían contribuir en la mejora misma de los sistemas individuales y proporcionaron nuevos conocimientos científicos, especialmente en el contexto del procesamiento estadístico de datos.

6. Descripción del corpus y de las herramientas de análisis

Un aspecto clave del proyecto TaraXÚ es el hecho de que fue desarrollado en un ámbito real, en el que se tuvieron en cuenta las necesidades reales de los socios de la industria de la traducción, en lugar de apoyarse en criterios puramente teóricos. Por ello consideraron los investigadores que el proyecto puede ser una base mucho más fiable para el progreso científico de los procedimientos actualmente disponibles y ampliamente usados («DFKI LT - TaraXÚ» 2015).

En el transcurso del proyecto se recogieron a partir del 2010 un volumen de datos de aproximadamente 300.000 palabras en diferentes combinaciones lingüísticas, que se analizaron en cuatro etapas diferentes («DFKI LT - TaraXÚ» 2015). Las salidas de traducción investigadas en este proyecto son textos originales traducidos (en ambas direcciones) con la ayuda de sistemas de traducción automática en las combinaciones lingüísticas alemán-inglés, alemán-francés y alemán-castellano. No se trata de textos completos sino de un corpus de segmentos que provienen de tres dominios: textos informativos tomados de corpora públicos de WMT (Callison-Burch et al. 2011), documentación técnica extraída de la documentación de OpenOffice (Tiedemann 2009) y datos particulares de un cliente propiedad de los socios del proyecto.

Las salidas generadas por estos los sistemas de traducción empleados fueron presentadas a los traductores profesionales para que estos llevaran a cabo la clasificación y posesición con una herramienta de evaluación basada en el navegador *Appraise* (Federmann 2010). Las tareas se definieron de la siguiente manera: primero se debía clasificar, con las cifras del 1 al 4, cada traducción teniendo en cuenta qué tanto se conservaba el significado del segmento original. Luego tenían que marcar los errores. Posteriormente venía la fase de selección y posesición: para cada frase original se debía seleccionar la traducción que fuese más fácil de poseer y luego realizar la posesición (Popovic et al. 2013).

	Fase 1	Fase 2	Fase 3	Fase 4
Combinación lingüística	alemán–inglés alemán<->español	alemán<->inglés alemán<->español alemán<->francés checo<->inglés	alemán–inglés alemán–chino	alemán<->inglés alemán<->español alemán<->francés

6. Descripción del corpus y de las herramientas de análisis

	Fase 1	Fase 2	Fase 3	Fase 4
Núm. de palabras	50.000	350.000	35.000	300.000
Tareas	- Categorización - Clasificación de errores nivel de la oración - Selección y posesición	- Categorización - Clasificación de errores nivel de las palabras - Selección y posesición - Posedición de todas las traducciones	- Categorización - Selección y posesición	- Categorización - Clasificación de la calidad de la traducción - Selección y posesición

Tabla 106: Fases del proyecto TaraXÚ

Para la traducción de los textos del alemán al castellano se utilizaron cuatro diferentes sistemas de traducción⁸: *Moses*, como sistema estadístico basado en frases, *Jane* un sistema estadístico jerárquico, el sistema comercial basado en reglas *Lucy*, y también otro sistema comercial basado en reglas RBMT. Además se tuvo a disposición una memoria de traducción de un cliente. Los sistemas fueron anonimizados para la evaluación.

La evaluación fue realizada por traductores profesionales bilingües que trabajan para los proveedores de servicios que participaron en el proyecto. Los evaluadores tenían como tarea escoger entre las propuestas aquella que fuera más adecuada para la posesición y luego corregirla, cambiándola lo menos posible, hasta obtener una traducción aceptable. Los poseedores no recibieron ninguna guía especial, pero fueron guiados por la estructura misma de la interfaz de usuario desarrollada para el proyecto.

Al finales del 2013 contactamos directamente con la firma Euroscript. Después de presentar el objetivo de nuestra tesis obtuvimos un extracto del corpus de TaraXÚ en el que se encontraban los segmentos que se tradujeron del alemán al castellano. Nos comprometimos a utilizar este extracto solo para fines investigativos y a no divulgar a terceros la información incluida. El día 11 de septiembre del 2013 firmamos un acuerdo de confidencialidad sobre la

⁸ En la última etapa se utilizaron sólo tres motores, dos estadísticos y uno basado en reglas

6. Descripción del corpus y de las herramientas de análisis

transferencia de los resultados de la evaluación del proyecto. El “Acuerdo con Euroscript” se podrá ver Anexo 1 en el CD-ROM.

6.1.1 Estructura del corpus extraído de TaraXÚ

Recibimos de la firma Euroscript dos archivos en formato *.xml. El primer archivo contenía las traducciones de los sistemas de traducción automática y las propuestas de la memoria de traducción sin posesición (translations.xml). Aquí tenemos que anotar que solo 357 segmentos tenían una propuesta de una memoria de traducción, mientras que todos los 940 segmentos tienen las propuestas de traducción de los cuatro sistemas de traducción automática. El segundo archivo que recibimos contenía la propuesta seleccionada y su posesición (posediting-results.xml). Los datos del archivo con las traducciones tenían etiquetas con las que se identificaba la siguiente información:

Nombre del campo	Descripción
column name="id"	Número de identificación interna en el corpus de TaraXÚ.
column name=doc_name	Nombre del subcorpus en el archivo proporcionado por Euroscript (veáse tabla en la página 219).
column name="target"	Traducción realizada por alguno de los sistemas que están marcados con la etiqueta "system_name".
column name="source"	Segmento original en alemán.
column name="system_name"	Aquí aparece RBMT1, RBMT2, SMT1, SMT2 o TM.

Tabla 107: Etiquetas encontradas en el archivo translations.xml

Los datos del archivo con las posesiciones tenían ocho etiquetas con las que se podían identificar los siguientes datos:

Nombre del campo	Descripción
column name="id"	Número de identificación interna en el corpus de TaraXÚ.
column name="src_sentence"	Segmento original en alemán.

6. Descripción del corpus y de las herramientas de análisis

Nombre del campo	Descripción
column name="tgt_sentence"	Traducción realizada por el sistema identificado con la etiqueta "system".
column name="duration"	Tiempo empleado para la posesición.
column name="user_id"	Número de identificación interna del poseedor (hubo dos evaluadores).
column name="postedited"	Segmento poseído.
column name="system"	Se indica el sistema cuya propuesta se escogió para la posesición.
column name="corpus_name"	Nombre del corpus de donde se extrajo el original.

Tabla 108: Etiquetas encontradas en el archivo posediting-results.xml

Según el origen de los textos que pudimos ordenar los datos con ayuda del campo “corpus_name” y pudimos identificar los siguientes subcorpus: los segmentos que pertenecen a “shared/guide” forman parte del corpus público de documentación de una suite ofimática libre («OpenOffice» 2009); tenemos también segmentos originalmente escritos en castellano, checo e inglés que hacen parte de corpus públicos de noticias. Además tenemos 1081 segmentos escritos originalmente en alemán y que pertenecen a socios del proyecto. Este último criterio es importante para la tesis, pues gracias a él se puede determinar si el texto ha sido originalmente redactado en alemán o si se trata de una traducción de otro idioma al alemán y luego al castellano. No solo el hecho de que los segmentos fueron escritos originalmente en alemán, sino además que son traducciones de manuales y guías de usuario fue el criterio decisivo que nos llevó a extraerlos y analizar con ellos los problemas y errores de traducción.

Nombre	Idioma origen	Número de segmentos
shared/guide	Inglés	419
customer	Alemán	1081
iDNES.cz	Checo	26
Aktulne.cz	Checo	23

6. Descripción del corpus y de las herramientas de análisis

Nombre	Idioma origen	Número de segmentos
Ihned.cz	Checo	30
Novinky.cz	Checo	40
frankfurter-rundschau (Frankfurter Rundschau)	Alemán	18
Spiegel (Der Spiegel)	Alemán	30
Faz (Frankfurter Allegemeine Zeitung)	Alemán	18
Sz (Süddeutsche Zeitung)	Alemán	18
Ftd (Financial Times Deutschland)	Alemán	18
economist (The Economist)	Inglés	24
latimes (Los Angeles Times)	Inglés	36
Nytimes (New York Times)	Inglés	24
washpost (Washington Post)	Inglés	24
Abces (ABC Español)	Castellano	36
Cinodias (Cinco días)	Castellano	33
Elperiodicoes (El periódico España)	Castellano	30
milenio	Castellano	7

Tabla 109: Lista de subcorpus encontrados

Con ayuda de las etiquetas y la información recibida por Euroscript pudimos extraer de este primer corpus solo los segmentos que fueron escritos originalmente en alemán, luego descartamos los segmentos de los subcorpus de noticias ya que no solo nos interesan para el análisis los segmentos que estén escritos originalmente en alemán, sino también aquellos cuyo tipo de texto es de orden técnico. Utilizamos para nuestra tesis el subcorpus marcado con la etiqueta <costumer>. Este subcorpus constaba de 1081 segmentos originales, con las traducciones de los cuatro sistemas empleados en la segunda y tercera fase del proyecto TaraXÚ. En algunos casos teníamos a disposición también las traducciones de una memoria de traducción.

6. Descripción del corpus y de las herramientas de análisis

Una vez tuvimos los segmentos del corpus <customer> eliminamos los segmentos que contenían solo números o cifras (salvo las ocasiones en que se hicieron cambios en la traducción y/o posesición) y las repeticiones; finalmente obtuvimos un total de 940 segmentos.

6.1.2 Reorganización de los datos del corpus

Con el fin de obtener una base homogénea y fácil de manejar para la investigación organizamos en primer lugar la información que recibimos de la firma Euroscript en los dos archivos *xml, cuya estructura describimos anteriormente, en una hoja de cálculo. En la tabla a continuación listamos los criterios con que ordenamos la información para iniciar el análisis. Frente a cada segmento original y cada propuesta de traducción (incluyendo las propuestas de la memoria de traducción) añadimos una columna para el conteo automático del número de palabras con ayuda de las funciones del programa de la hoja de cálculo. También contamos el número de palabras del segmento posesitado. Gracias a esta información pudimos calcular, con ayuda de la distancia de Levenshtein, la distancia de edición entre cada propuesta de traducción y la posesición. Este cálculo nos ayudó a identificar qué segmentos fueron posesitados y cuáles no, como describiremos en el capítulo de descripción del análisis del corpus.

Columna	Nombre	Descripción
Columna A	doc_name	En esta columna se encuentra el origen o nombre con que se marcó el segmento en el archivo *xml original. En nuestro caso aparece como “doc_name” solo “customer”. Esta columna la utilizamos en un primer momento para comprobar que el segmento procedía de este subcorpus y no de otro, por ejemplo de noticias.
Columna B	source	En esta columna se encuentran todos los segmentos originales en alemán.
Columna C	pal	Número de palabras del segmento original que se encuentra en la columna B.
Columna D	tm	Propuesta de la memoria de traducción, en caso tal de que hubiese una propuesta. De lo contrario se repite el segmento original.
Columna E	pal	Número de palabras de la propuesta dada por la memoria de traducción.
Columna F	Lv	Valor del cálculo con la distancia Levenshtein con respecto a la posesición.
Columna G	smt1	Propuesta de traducción automática del primer motor estadístico.
Columna H	pal	Número de palabras de la propuesta de traducción automática del primer motor estadístico que se encuentra en la columna G.
Columna I	Lv	Valor del cálculo con la distancia Levenshtein con respecto a la posesición.

6. Descripción del corpus y de las herramientas de análisis

Columna	Nombre	Descripción
Columna J	smt2	Propuesta de traducción automática del segundo motor estadístico.
Columna K	pal	Número de palabras de la propuesta de traducción automática del segundo motor estadístico que se encuentra en la columna J.
Columna L	Lv	Valor del cálculo con la distancia Levenshtein con respecto a la posesición.
Columna M	rbmt1	Propuesta de traducción automática del primer motor basado en reglas.
Columna N	pal	Número de palabras de la propuesta de traducción automática del primer motor basado en reglas que se encuentra en la columna M.
Columna O	Lv	Valor del cálculo con la distancia Levenshtein con respecto a la posesición.
Columna P	rbmt2	Propuesta de traducción automática del segundo motor basado en reglas.
Columna Q	pal	Número de palabras de la propuesta de traducción automática del segundo motor basado en reglas que se encuentra en la columna P.
Columna R	Lv	Valor del cálculo con la distancia Levenshtein con respecto a la posesición.
Columna S	postedited	Posedición de la propuesta seleccionada.
Columna T	pal	Número de palabras del segmento poseeditado, que se encuentra en la columna S.
Columna U	system	Nombre del sistema, cuya propuesta se escogió para la posesición.
Columna V	duration	Tiempo que necesito el poseedor para la posesición.

Tabla 110: Elementos empleados para la organización del corpus

En la siguiente ilustración podemos ver la una captura de pantalla de la tabla del archivo que contenía los 940 segmentos extraídos y reorganizados, donde se incluyen también las propuestas de la memoria de traducción:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1	doc_name	source	pal	tm	pal	Lv	smt1	pal	Lv	smt2	pal	Lv	rbmt1	pal	Lv	rbmt2	postedited	pal	system	duration	
146	customer	140.000	1	0	140.000	1	0	140.000	1	0	140.000	1	0	140.000	1	0	140.000	140.000	1	tm	00:00:04.3
147	customer	"<field name="Sparatext"/>" auf Si	6	62	"las cosas	21	72	"Sucedá f	18	67	"<field na	6	59	"<field	6	59	"<field	"<field name="\$par	7	rbmt1	00:00:40.8
148	customer	"<field name="Sparatext"/>" auf Si	6	62	"las cosas	21	72	"Sucedá f	18	67	"<field na	6	59	"<field	6	59	"<field	"<field name="\$par	7	rbmt1	00:00:20.4
149	customer	"Best in Class" Gesamtwirkungsgra	9	71	"mejor e	12	73	"« mejor e	14	67	"Best en	17	31	"Best en	17	31	"Best en	"<Best in Class" Rendim	17	rbmt1	00:02:11.1
150	customer	(Abschnitt <field name="Stblsheet	6	47	(Abschn	6	47	(sección	18	56	(Apartad	14	58	(Cortaba	6	47	(sección	(sección<field name="\$tb	4	rbmt2	00:02:10.4
151	customer	(Abschnitt <field name="Stblsheet	6	47	(Abschn	6	47	(sección	18	56	(Apartad	14	58	(Cortaba	6	47	(sección	(sección<field name="\$tb	4	rbmt2	00:00:16.3
152	customer	(Forts.)	1	0	(forts.)	1	4	(Forts.)	1	4	(Forts.)	1	4	(Fuertes,	1	6	(Forts.)	(contin.)	1	tm	00:00:04.6
153	customer	(wie z. B. Wilo-Stratos PICO und W	8	17	(como, p	8	1	(Como, p	8	6	(como a	8	10	(como W	8	10	(como por ejemplo Wilo-Stratos P	8	smt1	00:00:10.4	

Ilustración 27: Tabla de datos con conteo de palabras y de distancia de teclado

Como segundo paso en la organización del corpus comprobamos que en la columna A “doc_name” no se encontrara un segmento que no perteneciera el subcorpus <customer> y eliminamos la columna. También eliminamos la columna <duration> porque los valores que aquí se encontraban no daban ninguna información exacta: en algunos casos la duración tenía

6. Descripción del corpus y de las herramientas de análisis

un valor muy alto, a pesar de que no se realizó ningún cambio en la posesición o los cambios fueron mínimos.

Una vez elaborada esta tabla hicimos un primer análisis de los datos contando el número de palabras de cada segmento (original, propuesta de traducción y posesición) para poder así calcular la distancia de Levenshtein. Aquí pudimos identificar qué propuestas de los sistemas de traducción fueron consideradas como correctas por los poseedores: si no se realiza ningún cambio, es decir si la propuesta del motor de traducción automática no es cambiada durante la posesición, el valor será 0, si se realizan muchos cambios el valor es mayor que si solo se realizan pocas modificaciones.

Para complementar el análisis puramente cuantitativo de los datos que teníamos elaboramos otra hoja de cálculo, en la que no tuvimos en cuenta la distancia de teclado, ni el número de palabras. Allí añadimos una columna para describir de forma cualitativa las ediciones realizadas durante la posesición en el segmento traducido y, si es reconocible, el origen de las ediciones en el segmento de partida. Marcamos con color azul la propuesta que fue escogida para la posesición. Como indicaremos en los apartados sobre cada fase en detalle, el primer análisis de datos los realizamos con el corpus de 940 segmentos y todas las propuestas de traducción realizadas. Para el análisis detallado de los errores que corrigieron los poseedores utilizaremos solo las propuestas que fueron seleccionadas para la posesición y que no eran un *fuzzy-match* de una memoria de traducción. Allí tuvimos un subcorpus de 650 segmentos, con la propuesta del sistema de traducción seleccionada por el poseedor y la respectiva posesición.

Tal como indicamos en la introducción de esta parte (véase página 202), en nuestro análisis realizamos una descripción cuantitativa y cualitativa de los errores encontrados en las propuestas de traducción con el fin de elaborar una clasificación que pueda eventualmente conducir a formular criterios para la evaluación de traducciones automáticas, y sobre todo para crear directrices que ayuden en el proceso de posesición.

Realizamos el procesamiento del corpus en tres fases que describiremos detalladamente en apartados separados. La primera fase no fue un intento completamente válido de clasificación, sino más bien un punto de partida para las siguientes fases. Esta fase consistió en un

6. Descripción del corpus y de las herramientas de análisis

análisis cuantitativo detallado de los errores encontrados y una clasificación a grandes rasgos. En la segunda fase del análisis establecimos las categorías y tipos de errores con ayuda de la propuesta de errores que hicimos a partir del marco teórico y de los errores que encontramos en la primera fase. En la tercera fase clasificamos los errores encontrados según una tipología de MQM, y con el fin de constatar si una categorización diferente nos da la posibilidad de especificar los errores encontrados de una manera más adecuada o si hay tipos de errores que no tuvimos en cuenta en nuestra propuesta.

Para las dos primeras fases del análisis utilizamos un archivo de cálculo diferente y en el archivo mismo empleamos diferentes hojas. En el archivo que denominamos primer análisis le dimos a los datos del corpus un orden diferente, que facilitara la anotación y el conteo de errores. En la primera hoja del archivo encontramos cuatro columnas a la que no les dimos títulos:

Columna	Nombre	Descripción
Columna A	Número	Número de identificación del segmento de 001 a 940 en nuestro corpus.
Columna B	Sistema	Nombre del sistema que con el que se elaboró la propuesta que se encuentra en la columna C. Los valores que podemos encontrar aquí son: source, tm, smt1, smt2, rbmt1, rbmt2 y postedited.
Columna C	Original o traducción	Aquí se encuentran —según lo que aparezca en la columna B— el segmento original, la propuesta de la memoria de traducción o la traducción automática de un motor en particular y la posesición.
Columna D	Anotaciones	Aquí se realizaron las anotaciones, es decir, se describieron de los errores encontrados y también anotamos si la propuesta era correcta o si no había propuesta.
Columna E	Tipo de error	En esta columna aparece una denominación general del error que utilizamos para el conteo de errores. También, si es el caso, aparece si la propuesta es correcta.

Tabla 111: Descripción de la tabla para las anotaciones en la primera fase

La segunda hoja la utilizamos para listar los tipos de errores encontrados y para el recuento de los mismos. En hojas separadas hicimos las gráficas de los errores que se encontraron más frecuentemente en cada sistema de traducción, incluyendo los errores que encontramos en las propuestas de la memoria de traducción.

6. Descripción del corpus y de las herramientas de análisis

Para el segundo análisis partimos del mismo archivo de cálculo que utilizamos para el primer análisis, pero aquí analizamos solo los errores encontrados en las propuestas que se seleccionaron para la posesición y no en todas las propuestas de traducción. Además eliminados los segmentos en los que se escogieron las propuestas de la traducción de memoria. En la primera hoja hicimos, como durante el primer análisis, la anotación de los errores y en la segunda hoja el listado de las categorías y el conteo de los errores encontrados. En esta fase no anotamos los errores de manera general, sino que utilizamos una clasificación que explicaremos más adelante (véase página 283).

En la tercera fase utilizamos las categorías de errores de MQM utilizamos el mismo corpus que empleamos en la fase dos, pero la clasificación y el conteo de errores los hicimos con ayuda de dos herramientas desarrolladas por el grupo de investigación QTLaunchPad del DFKI (*Deutsches Forschungszentrum für Künstliche Intelligenz*) que describiremos en el siguiente apartado a partir de la página 224: los árboles de toma de decisiones y el programa *translate5*. La finalidad del proyecto fue desarrollar un sistema en línea que ofreciera un fácil acceso a los recursos y las herramientas relacionadas con la traducción automática y que facilitará la comunicación entre los desarrolladores de los sistemas de traducción automática y los expertos en traducción. En el proyecto desarrolló también la métrica de evaluación de la calidad MQM, y un programa para la estimación de calidad que proporcionan una estimación de la calidad de una traducción sin hacer uso de traducciones de referencia. El investigador principal del proyecto fue Hans Uszkoreit. El proyecto fue una iniciativa financiada por la Unión Europea y tuvo una duración de 24 meses, desde el primero de junio de 2012 a junio del 2014.

Los resultados que obtuvimos en cada análisis se presentarán los capítulos correspondientes. Antes de entrar en detalle presentaremos las herramientas que utilizamos para realizar la comparación de los resultados obtenidos en el análisis descriptivo.

6.2 *Herramientas adicionales utilizadas*

Como indicamos anteriormente para la evaluación de nuestro corpus con la tipología de errores de MQM utilizaremos dos herramientas desarrolladas para la investigación del grupo QTLaunchPad. Nos referimos en primer lugar a los *Decision Trees* y en segundo lugar a la

6. Descripción del corpus y de las herramientas de análisis

herramienta basada en web de código abierto desarrollada para la corrección de pruebas, evaluación y posesición de traducciones *translate5* (Mittag 2013).

6.2.1 Árboles de decisiones

Los árboles de toma decisiones son herramientas empleadas para el aprendizaje de cómo utilizar la categorización de errores de MQM y cómo hacer una distinción clara entre ellos, ya que ayudan en la determinación en los casos en que la respuesta no es inmediatamente evidente. Para calificar o marcar los errores en la tercera fase nos guiamos por el árbol de decisiones del QTLaunch Pad (Burchardt y Lommel 2014).

El modelo de del árbol de decisiones trabajado por los investigadores del proyecto mencionado anteriormente está diseñado para que con una batería de preguntas se pueda identificar claramente cada tipo de error. Es decir, se formulan preguntas con respuestas binarias (o si son más de dos, siempre limitadas), y a cada respuesta corresponde o una categoría o una nueva pregunta, también binaria. Este requisito es importante porque todos los elementos o categorías de error derivados de un tipo de tema en particular (un problema y sus derivados constituyen una rama) podrían ser clasificados erróneamente como error de tipo general; por ello es necesario que con una sola pregunta se pueda hacer una distinción clara entre cada categoría y subcategoría de error.

Después de determinar qué error está contenido en cada nodo, es importante definir los tipos de error más específicos antes de los más generales. Este tipo de directriz funciona según el principio de la exclusión: mediante la eliminación de los casos específicos se llega al caso general. Por ejemplo en el caso de la categoría general *Accuracy* primero debemos determinar si se trata de una traducción errónea (*Mistranslation*) o si no si se trata de una omisión (*Omission*). Dado que estos son tipos de error específicos de *Accuracy*, tienen que ser eliminados antes de declarar el error como un problema general de exactitud. Lo que quiere decir, que no solo tenemos la posibilidad de especificar exactamente la subcategoría a la que pertenece el error, también se puede clasificar un error como un fallo general en la fluidez, que no se puede clasificar bajo ninguna de las subcategorías propuestas por la tipología de MQM.

6. Descripción del corpus y de las herramientas de análisis

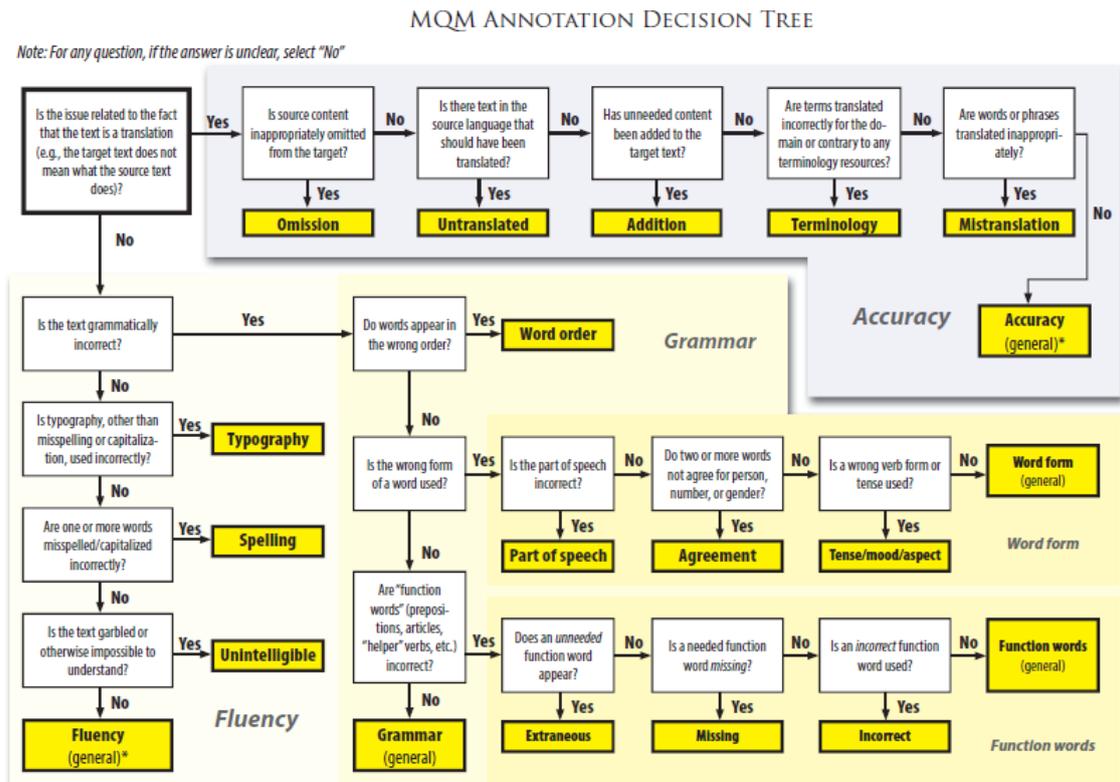
De acuerdo con lo anterior, un árbol de toma de decisiones para identificar los errores que se dan en la categoría *Accuracy* según la definición de MQM puede ser como describiremos a continuación. Los autores recomiendan en caso de duda optar por la respuesta negativa (Burchardt y Lommel 2014, 10-12):

- ¿Se ha omitido en la traducción algún contenido presente en el original?
- Si se contesta afirmativamente vaya a la pregunta 2 [se sabe que es un error de la subcategoría *Omission*]
- Si se contesta negativamente, vaya a la pregunta 3 [Se ha excluido *Omission* y ahora se debe tratar de definir de qué tipo subcategoría de *Accuracy* se trata].
- ¿Se ha omitido una variable en la traducción? [se trata de determinar el subtipo específico de la categoría *Omission*].
- Si se contesta afirmativamente: el sub-tipo de error es variable omitida.
- Si se contesta negativamente se ha excluido el subtipo de omisión, dejando el tipo general *Omission*.
- ¿Se han traducido palabras o frases erróneamente? [está pregunta selecciona o excluye la categoría *Mistranslation*]
- Si se contesta afirmativamente: ir a la pregunta 4 [el error es de este tipo]
- Si se contesta negativamente [se han excluido *Omission* y *Mistranslation* dejando solo *Accuracy*].
- ¿Se han traducido erróneamente los números? [seleccionar o excluir el subtipo *Number*]
- Si se contesta afirmativamente: error del subtipo *Number*
- Si se contesta negativamente ir a la pregunta 5 [se ha excluido, debe ser otro tipo de error].
- ¿Se ha traducido erróneamente una palabra o una frase de dominio u organización específicos? [está pregunta selecciona o excluye la categoría *Mistranslation*]
- Si se contesta afirmativamente: ir a la pregunta 6 [sabemos que se trata de un error de terminología]
- Si se contesta negativamente: *Mistranslation* [se excluye cualquier otra opción]

6. Descripción del corpus y de las herramientas de análisis

- ¿Se ha traducido la palabra o frase contra de las directrices de terminología específicos de la empresa? [esta pregunta selecciona o excluye la categoría *Company Terminology*]
- Si se contesta afirmativamente, se trata de este subtipo de error
- Si se contesta negativamente, se excluye la subcategoría y queda la categoría general *Terminology*.

Aunque el árbol de decisiones completo de MQM (versión de noviembre de 2014, contiene 115 tipos de errores categorizados en cinco ramas principales) no está destinado a ser utilizado en su totalidad para cualquier tarea de evaluación; en la tabla resumen que vemos a continuación se presenta un “árbol de decisiones” general. En términos prácticos los autores recomiendan, utilizar un árbol de decisiones más pequeño de acuerdo con el proyecto. Para utilizar el árbol de decisiones hay que empezar con la primera pregunta y seguir las respuestas apropiadas hasta que se alcance un tipo de problema específico:



* Please describe any **Fluency (general)** or **Accuracy (general)** issues using the Notes feature.

Ilustración 28: Ejemplo de un árbol de decisiones

6. Descripción del corpus y de las herramientas de análisis

6.2.2 *translate5*

Como indicamos al comienzo esta herramienta ha sido desarrollada para el QTLAuch por la firma MittagQI (Mittag 2013). Para poder trabajar con esta herramienta tuvimos que convertir la tabla Microsoft Excel ® de nuestro corpus en un archivo *csv que fue trabajado según las especificaciones de formato de QTLAuch Pad (Hans Uszkoreit 2014).

Esta herramienta nos permite evaluar las traducciones según los criterios que se seleccionen, si no se hace ningún tipo de adaptación se tendrá a disposición una versión reducida de la tipología de errores de MQM. Nosotros optamos por utilizar todos los tipos de errores de la definición (actual) de MQM. Para ello tuvimos que hacer un archivo en formato *xml en el que se declaraban todos las categorías y subcategorías de errores de MQM (véase anexo en el CD-ROM).

En la interface de *translate5* podíamos ver el segmento original y el segmento que fue escogido para la posesición. Al lado del campo con el segmento seleccionado para la posesición se encuentra el campo para la marcación de los errores. No solo se pueden marcar los tipos de errores, también teníamos la opción de darles un peso: *minor*, *major* y *critical*, pero no utilizaremos esta opción en el análisis del corpus.

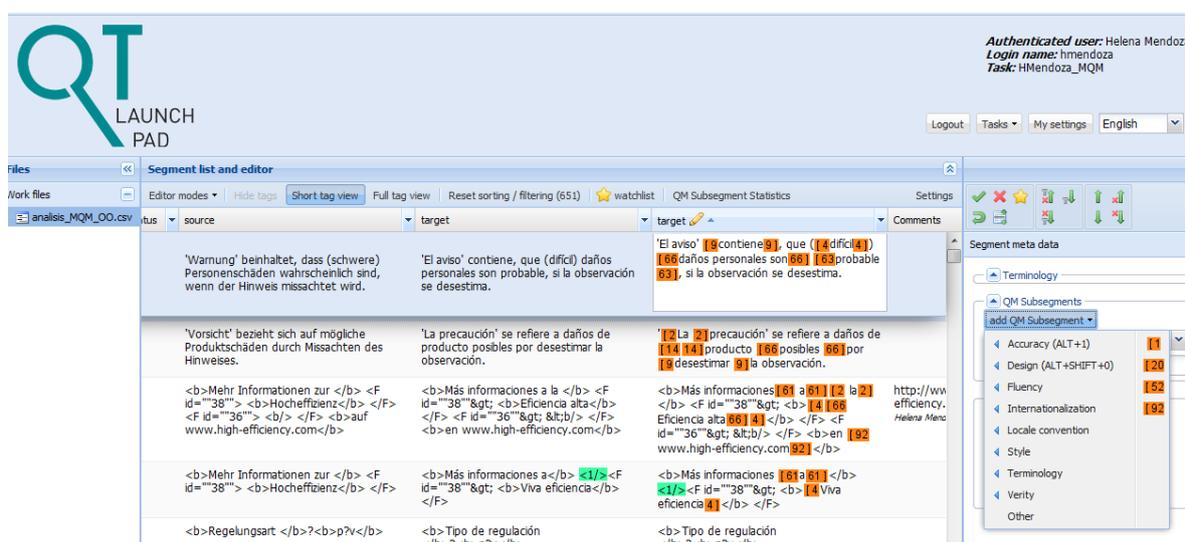


Ilustración 29: Interface de usuario de translate5

Al finalizar la clasificación de los errores según la métrica de MQM se obtiene una tabla estadística con todas las marcaciones.

6. Descripción del corpus y de las herramientas de análisis

Para analizar los datos según las categorías de MQM en translate5 tuvimos que crear un archivo en formato *xml en el que se encontraban el número de identificación del segmento, el original y la propuesta posestada, organizados con las etiquetas "mid", "source" y "target" respectivamente. Este archivo fue estructurado y preparado según las instrucciones para la utilización del programa.

También tuvimos que crear un archivo en *xml donde se encontrarán los tipos de errores de la clasificación de MQM que queríamos utilizar. Decidimos utilizar todas las categorías y subcategorías y no excluir ninguna. Una vez cargamos en MQM nuestro proyecto, es decir los segmentos que fueron escogidos para la posesición y las categorías de errores de MQM, realizamos la marcación de errores en la interfaz de translate5 y al finalizar las anotaciones pudimos exportar los resultados en un archivo *csv. Con base en los datos guardados pudimos establecer el total de errores en las categorías de MQM de primer rango y en cada subcategoría:

MQM Core	Categorías primer rango	Categorías segundo rango	Categorías tercer rango	Número de errores	Total por categorías
Accuracy	Addition			191	191
	Improper-exact-tm-match				
	Mistranslation			154	506
		Ambiguous-translation			
		Date/time			
		Entity		14	
		False friend			
	Mistranslation of technical relationship			288	

Ilustración 30: Tabla de conteo de errores según MQM

La ilustración anterior muestra la tabla de conteo de errores de la categoría *Accuracy* con las categorías de primer y segundo rango y el número de errores marcado en cada uno de ellas.

En la categoría *Accuracy* encontramos la mayor cantidad de errores, seguida de la categoría *Fluency* y luego de *Terminology*. En la ilustración a continuación podemos ver el porcentaje de errores que contenían las traducciones automáticas en bruto seleccionadas por los poseedores:

6. Descripción del corpus y de las herramientas de análisis

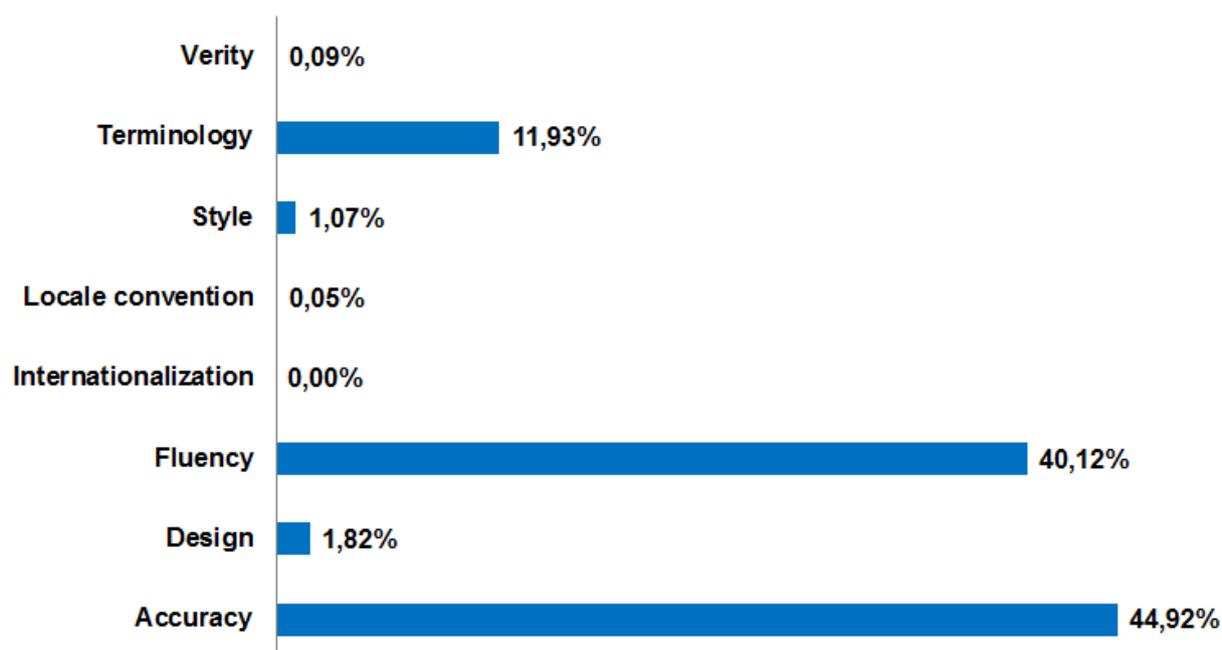


Ilustración 31: Errores encontrados en las categorías generales de MQM

En la tabla a continuación podemos ver el número de errores marcados en cada categoría del MQM Core:

Categoría	Total de errores marcados
<i>Accuracy</i>	964
<i>Design</i>	39
<i>Fluency</i>	861
<i>Internationalization</i>	0
<i>Locale convention</i>	1
<i>Style</i>	23
<i>Terminology</i>	256
<i>Verity</i>	2
TOTAL	2146

Tabla 112: Errores marcados en las macrocategorías de MQM

No marcamos ningún error en la categoría *Internationalization* porque esta categoría ha sido desarrollada especialmente para la evaluación de la localización de contenidos. Si bien los errores de esta categoría se detectan generalmente en el texto de destino, se refieren a pro-

6. Descripción del corpus y de las herramientas de análisis

blemas en los originales que causan problemas en el contenido traducido/localizado. Por ejemplo, si se presentan errores funcionales por la falta de apoyo para los idiomas como guiones, signos de interrogación, comillas u otras características específicas del entorno local. En nuestro corpus no teníamos la información necesaria que nos permitiera evaluar este tipo de errores. A continuación daremos un listado de los errores marcados en cada categoría con ejemplos para ilustrarlos.

7. Análisis preliminar del corpus y descripción de los motores de traducción

En este capítulo haremos una descripción detallada del análisis del corpus partiendo de los datos que enunciamos en el apartado anterior. Nuestro propósito es abordar de forma sistemática los problemas con que se han encontrado en la posesición y analizar qué correcciones ha realizado el poseedor de forma descriptiva. Nuestro análisis se divide en tres fases: Empezaremos con el análisis descriptivo de cómo actuaron los motores y de qué modo enfrentaron los poseedores los problemas y errores encontrados. El segundo análisis está más orientado a detectar qué tipo de problemas se solucionan, de qué manera y con qué estrategia, luego describiremos los resultados del análisis del corpus según la métrica de MQM y finalmente haremos una comparación de los resultados de los dos análisis.

No corregiremos ni analizaremos los errores que no hayan sido detectados en la posesición y solo nos centraremos en los que el poseedor corrigió, así la corrección no sea la mejor o no sea completamente correcta.

En la primera fase del estudio analizamos todas las propuestas de traducción y no solo las propuestas que se escogieron para la posesición. De este modo conseguimos un acercamiento mucho más profundo a la naturaleza de los datos del corpus. Anotamos todos los errores de cada propuesta de traducción automática detalladamente. Una vez realizamos todas las anotaciones pudimos establecer ciertos tipos de errores que se repetían. A partir de la descripción detallada llegamos a un listado más esquemático de las categorías. Finalmente, hicimos en la hoja dos un conteo de los errores encontrados en cada propuesta de traducción, en las propuestas de la memoria de traducción y también en los segmentos originales y en las posesiciones. En un primer momento tuvimos una lista provisional de descripciones de errores de más de 400 registros (ver Anexo 3: Análisis primera fase); después de analizar la lista redujimos eliminando las redundancias y solapamientos a 234 tipos de errores (ver página 238). Estos a su vez pueden ser agrupados en categorías más generales.

7. Análisis preliminar del corpus y descripción de los motores de traducción

A	B	C	D	
Número	Sistema		Comentarios	Tipo de error
001	source	14.000		
	tm	14.000	correcta	correcta
	smt1	Los 14 000	se añade innecesariamente el artículo	se añaden palabras
	smt2	14,000	se cambia punto por coma	error de convenciones
	rbmt1	14.000	correcta	correcta
	rbmt2	14.000	correcta	correcta
	postedited	14.000		
	002	source	"<field name="\$paratext"/>"	

Ilustración 32: Tabla para la anotación de los errores

Ya que la primera fase se basa en describir cómo actuaron los motores y de qué modo se enfrentaron los poseedores, en este primer análisis pudimos determinar en primer lugar qué tipo de motor fue favorecido por los poseedores y en qué motores encontramos más errores. Además, siguiendo el análisis cuantitativo, pudimos determinar qué tipos de errores se presentaban en cada sistema de traducción automática y en qué cantidad.

Según los datos proporcionados por Euroscript pudimos establecer también cuántos segmentos se positaron a partir de cada sistema de traducción automática. En este conteo obtuvimos como resultado que los poseedores seleccionaron en 289 ocasiones las propuestas de la memoria de traducción, lo que supone el 30% de los segmentos del corpus. En 111 casos las propuestas del SMT1 y en 78 las del SMT2. El sistema cuyas propuestas fueron más favorecidas por los poseedores fue el RBMT2 (413) y el menos favorecido el RBMT1, cuyas propuestas solo fueron escogidas en 49 ocasiones. Hay que tener en cuenta también que mientras para cada segmento original teníamos las cuatro propuestas de los motores de traducción automática, solo 357 de los 940 segmentos tenían una propuesta de la memoria de traducción.

7. Análisis preliminar del corpus y descripción de los motores de traducción

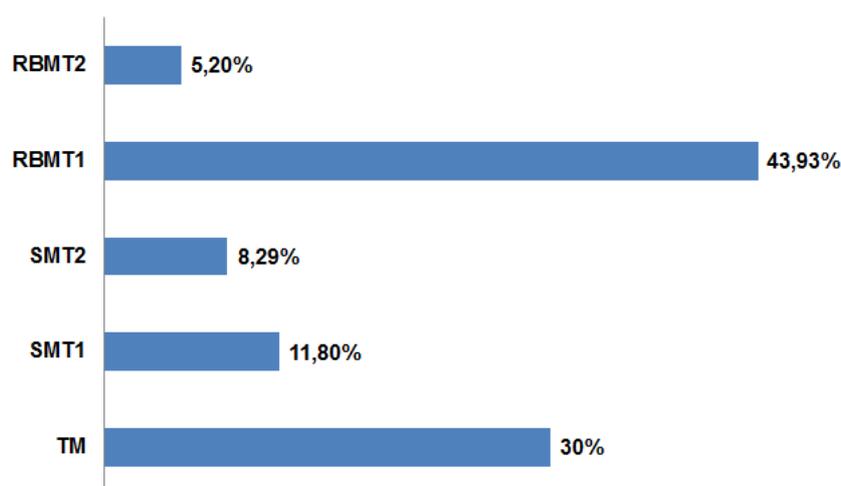


Ilustración 33: Segmentos posestados a partir de cada sistema (incluyendo las propuestas de la TM)

7.1 Categorización de los errores

Después de analizar las primeras anotaciones y datos agrupamos los diversos tipos de errores en diferentes categorías. Para ello dividimos el análisis en dos hojas de cálculo diferentes. En la segunda hoja del archivo de cálculo hicimos una tabla que contenía las siguientes columnas: en la primera columna (o columna A) listamos las categorías de problemas en orden alfabético y al frente de ellas en cada columna el número de errores encontrados en cada uno de los sistemas, en los originales o en las posestaciones, es decir: source, postedited, tm, SMT1, SMT 2, RBMT1 y RBMT 2 (columna B a H). Es decir no solo anotamos los errores encontrados en las propuestas de la memoria de traducción y en las posestaciones, sino que también hicimos un conteo de ellos. Los datos procesados se encuentran en el Anexo 3: Análisis primera fase.

	A	B	C	D	E	F	G	H
	Categoría de error	source	postedited	tm	smt1	smt2	rbmt1	rbmt2
1	correcta			209	61	64	198	97
2	error de convenciones		3		5	3	3	1
3	error de estilo		13	5	7	5	9	2
4	error de ortografía			1	3	5	1	7
5	error de puntuación		1		16	30	13	13
6	error de puntuación (falta signo)		9	18	33	42	18	36
7	error de puntuación (posición errónea)				17	21	5	20
8	error de puntuación (se añade un signo)		2	3	62	82	17	50
9	error de puntuación (signo erróneo)		2	3	1	2	1	2
10								

Ilustración 34: Hoja de cálculo para conteo de errores según motor y tipo

7. Análisis preliminar del corpus y descripción de los motores de traducción

En las propuestas de la memoria de traducción encontramos en total 333 errores. En las traducciones automáticas marcamos en total 17.793 errores. Las propuestas con menos errores fueron las realizadas por el RBMT1 (3081 errores). Luego las propuestas de RBMT2 (4436 errores). Las propuestas con más errores fueron las realizadas con los motores estadísticos: SMT1 (5090 errores) y SMT2 (5072 errores). También encontramos 113 errores o problemas en los segmentos originales y 190 errores en las posesiciones. Estos últimos errores son marcaciones o correcciones nuestras. En la ilustración a continuación podemos ver el porcentaje de los errores marcados en las traducciones automáticas según el tipo de motor:

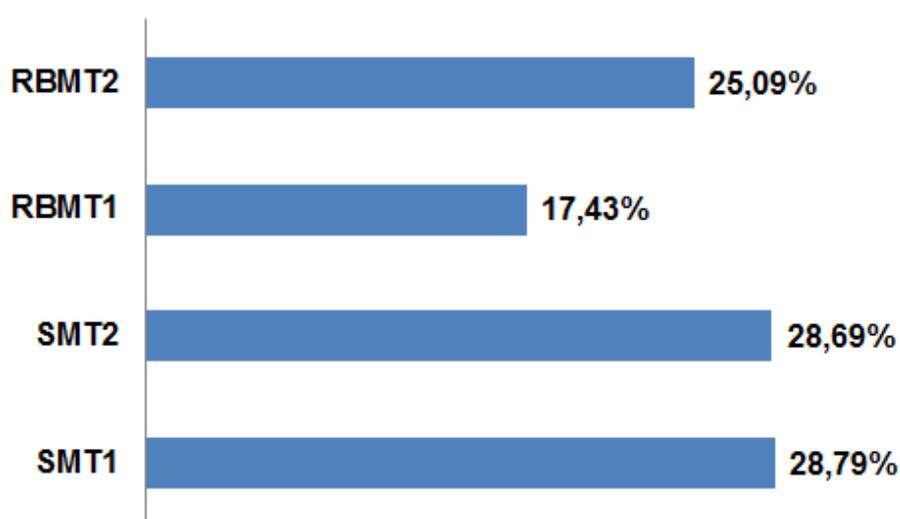


Ilustración 35: Porcentaje de errores según motor empleado para la traducción

Adicionalmente pudimos calcular la cantidad de propuestas correctas según el motor de traducción. Como propuestas correctas marcamos aquellas en las que no se tuvo que hacer ninguna posesición. Las propuestas que fueron extraídas de la memoria de traducción fueron generalmente aceptadas y no modificadas por los poseedores, por ello ocupan el primer lugar (209 propuestas correctas). En segundo lugar, encontramos las propuestas del motor de traducción RBMT1 (198 propuestas correctas), seguidas por las del RBMT2 (97 propuestas correctas). También marcamos como correctas las propuestas de traducción de los motores SMT2 (64 propuestas correctas) y SMT1 (61 propuestas correctas).

7. Análisis preliminar del corpus y descripción de los motores de traducción

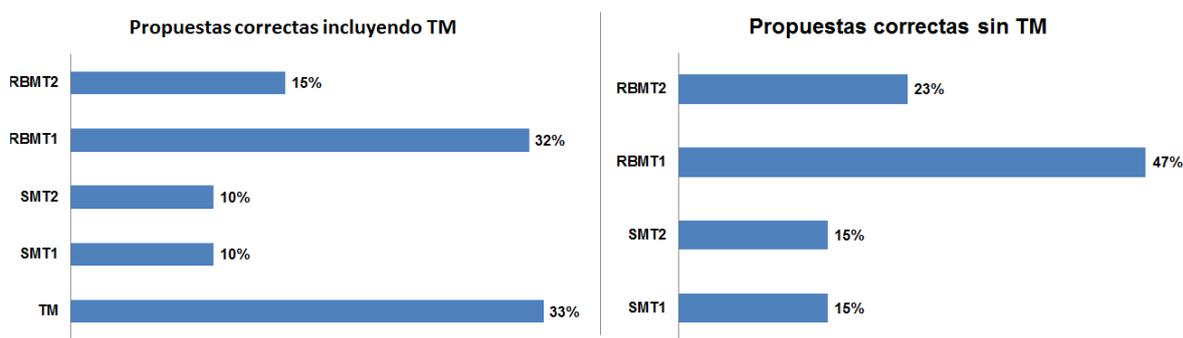


Ilustración 36: Propuestas marcadas como correctas según el motor

Además, hallamos también segmentos que no se tradujeron, es decir en los que no hubo propuesta: 583 segmentos no tuvieron traducción en la TM, 36 no fueron traducidos por el RBMT2, 21 por el SMT1, 20 por el SMT2 y solo 8 no fueron traducidos por el RBMT1. Hay que destacar que, si bien el 33% de las propuestas correctas provienen de un sistema de memoria de traducción, el 67% restante son traducciones automáticas. Aunque ningún sistema por su parte es equiparable a la cantidad de propuestas marcadas como correctas provenientes de la memoria de traducción (TM), todos los sistemas en conjunto sí la superan.

En los siguientes apartados describiremos los problemas y errores que hemos identificado en el corpus. En primer lugar, comenzaremos por los problemas provocados por las características del texto original. A continuación, abordaremos los que aparecen directamente en el segmento traducido.

7.2 Problemas y errores encontrados en los segmentos originales

En primer lugar, describiremos los errores o problemas que encontramos en los segmentos originales. Es decir, nos referimos no solo a faltas sino también tendremos en cuenta algunos aspectos, que a pesar de no ser errores pueden ser —como vimos en el marco teórico— problemáticos para la traducción automática en nuestra combinación lingüística.

Encontramos nueve tipos de errores o puntos problemáticos que listaremos en la tabla a continuación, en la que también daremos ejemplos e indicaremos cuántos errores se encontraron en cada una de las categorías. Solo daremos el segmento original no su traducción ni la posesición. Los errores serán marcados en color rojo para facilitar su identificación:

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Total	Ejemplo
Utilización de anglicismos	44	<i>Dies ermöglicht eine präzisere Steuerung der pump down, da die Sicherheits und Verzögerungszeiten bei dieser Bedienung eingestellt werden können.</i>
Puntuación errónea	22	<i>Eingang Niederdruck<ut/></i>
Errores formales	18	<i>Nachdem Sie die schematische Darstellung erstellt haben, können Sie diese über die Option Übersichtlicher einsehen einsehen einsehen oder über die Option Bearbeiten verändern.</i>
Nombres propios compuestos	11	<i>Wilo-Stratos und Wilo-Stratos PICO erfüllen jetzt schon alle Anforderungen von 2015</i>
Errores ortotipográficos	9	<i>In diesen Fällen ist keine spezielle Konfiguration not-wendig, es müssen lediglich die Parameter eingestellt werden...</i>
Uso incorrecto de mayúsculas/minúsculas	4	<i>Jetzt schon auf Hocheffiziente⁹ Produkte setzen.</i>

Tabla 113: Errores y puntos problemáticos más frecuentes hallados en los originales

Observamos que la utilización de anglicismos y otros términos propios del tipo de texto técnico y de la temática en particular —sin un entrenamiento especial del motor— pueden generar errores en la traducción automática. El problema de la utilización de anglicismos se da porque en algunos casos deben ser traducidos y en otros no. En segundo lugar, encontramos errores de puntuación, principalmente causados por la omisión de signos de puntuación. En tercer lugar, hallamos errores de tipo formal, como la falta de utilización de comillas u otro tipo de marcas para distinguir nombres de botones o comandos a dar durante la ejecución o instalación de un programa o para la delimitación de rutas o enlaces. Además, encontramos errores en la traducción de nombres propios, principalmente porque no se marcaron como tales o porque son compuestos que incluyen anglicismos o cifras. En algunos casos se combinan en los nombres propios diferentes unidades de sentido a modo de compuesto, como por ejemplo *Wilo-ErP-ready-Produkte*. En este caso “Wilo” es la marca registrada del fabricante,

⁹ En este caso el adjetivo no debe ir en mayúscula.

7. Análisis preliminar del corpus y descripción de los motores de traducción

“ErP” la sigla de una directiva de la Unión Europea, *ready* un anglicismo y *Produkte* el sustantivo alemán al que se refieren los otros componentes.

Finalmente encontramos errores ortotipográficos y algunas fallas e incoherencias en el uso de mayúsculas y minúsculas. En último lugar encontramos tres segmentos cuyo sentido era confuso, un error de estilo y un segmento en el que faltaba una preposición. En la ilustración a continuación podemos ver la proporción en que se presentan los seis tipos de errores más frecuentes que ya mencionados.

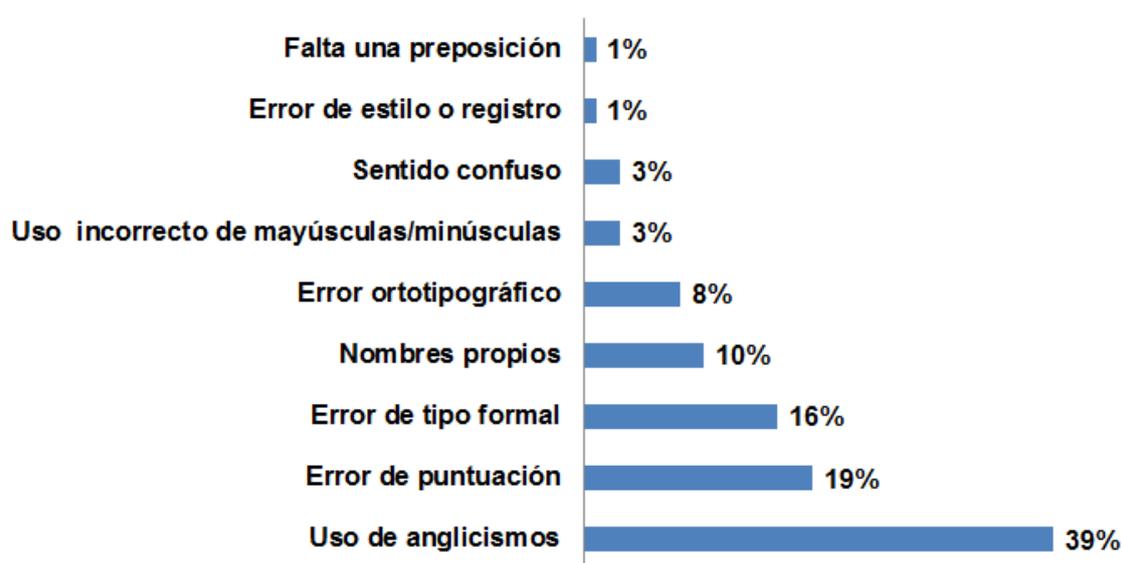


Ilustración 37: Problemas/errores marcados en los segmentos originales

7.3 Tipos de errores encontrados en las traducciones

Frente a cada propuesta de traducción anotamos los errores encontrados y así pudimos hacer un conteo tanto de los errores que más se presentaron de manera general y también filtrarlos teniendo en cuenta el motor empleado. Primero nos referiremos a los errores más frecuentes, sin referirnos al motor con que se realizaron las propuestas.

7.3.1 Descripción global de los errores en las traducciones

En esta primera clasificación encontramos 222 tipos de errores entre los que podemos listar errores de estilo, de ortografía, de puntuación, también errores formales y errores en el uso de convenciones, tanto del tipo de texto como terminológicas o culturales. Como errores de estilo marcamos aquellos que utilizan un registro inapropiado.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Los errores más frecuentes —sin tener el sistema de traducción automática empleado— son los errores ocasionados por la falta de coherencia o equivalencia, es decir, aquellos que se presentan cuando no se reproduce el sentido correctamente y hallamos construcciones que en castellano no generan sentido. Este tipo de errores son difíciles de clasificar ya que los problemas no resueltos que reflejan son diversos, pero tienen algo en común: impiden la comprensión del mensaje original. Es decir, en la mayoría de los casos se encuentran en un mismo segmento diversos tipos de errores, que en conjunto ocasionan el que se pierda el sentido.

En segundo lugar, encontramos que no se tradujo correctamente el sustantivo, no porque la traducción fuese en sí errada, sino porque no tiene en cuenta la terminología del tipo de texto. En tercer lugar, encontramos que el orden sintáctico de los elementos de la oración no era correcto; en cuarto lugar hallamos segmentos en los que no se tradujo el sustantivo, y en quinto lugar segmentos en los que se añadieron palabras que no se encontraban en el original. También hallamos traducciones confusas y algunas en las que la traducción del sustantivo o el verbo no se ajustaban al contexto.

Adicionalmente contamos como errores de tratamiento de las etiquetas aquellos casos en los que al ser interpretadas como texto se dañan y pierden su función; hallamos errores de puntuación, particularmente la omisión de signos de puntuación y también casos en que se reproduce la puntuación del original en alemán, sin ajustarla a las reglas de puntuación del castellano. Bajo errores formales clasificamos aquellos que tienen que ver con la demarcación errónea de rutas, la traducción de unidades de medida, cantidades, signos o números. También calificamos como un error formal los casos en que se colocaron erróneamente símbolos o la omisión de espacios p.ej. antes de un paréntesis.

Encontramos varios casos en que la traducción de las preposiciones en principio no era errónea pero no era la indicada en el idioma meta, ya que dependiendo de la función que tengan en la oración en alemán, en castellano debía usarse una preposición diferente. También se presentaron problemas al tratar de traducir la negación, en algunos casos por la posición del adverbio o de la conjunción ‘ni’, lo que en ciertos casos cambió el sentido mismo de la negación. Incluso encontramos un caso en que se incluyó una negación que no existía en el original.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Encontramos errores en la traducción de expresiones o colocaciones, que en algunos casos no se identificaron como tales y por ello no se tradujeron correctamente, por ejemplo, porque no se traducen todos los elementos, porque se añaden palabras o se utiliza la preposición incorrecta en castellano. En la tabla a continuación veremos algunos ejemplos de estos errores que marcamos en color rojo y la cantidad de errores que contamos en cada categoría. Para hacer más claro el error están marcados en color verde en el original las palabras que se tradujeron erróneamente y en la posedición la corrección. Algunos de los ejemplos presentan otro tipo de errores, pero solo marcamos el error que queremos ejemplarizar con ellos:

Tipo de error	Total	Ejemplo
Orden sintáctico erróneo de los elementos en la oración	1954	[TA]: Acceder a todo el mundo de producto de Wilo con la Wilo-App de forma en todo momento móvil [Original]: <i>Mit der Wilo-App jederzeit mobil auf die ganze Wilo-Produkt-Welt zugreifen</i> [PE]: Acceder con la Wilo-App en todo momento de forma móvil al mundo de los productos Wilo
Error global de sentido (falta de coherencia)	1724	[TA]: Para la eliminación del producto, y parte de ellas , la gestión pública o privada de las sociedades . ¹⁰ [Original]: <i>Zur Entsorgung des Produktes, sowie Teile davon, die öffentlichen oder privaten Entsorgungsgesellschaften in Anspruch nehmen.</i> [PE]: Para eliminar el producto o partes de éste , sírvase de empresas de eliminación de desechos públicas o privadas .
Elección léxica incorrecta del sustantivo (terminología)	1165	[TA]: Rebajamiento nocturno : [Original]: <i>Nachtabsenkung</i> : PE: Reducción nocturna :
Falta de traducción del sustantivo	945	[TA]: Los trabajos en el Produkt /der realización pueden ser realizados sólo en la parada. [Original]: <i>Die Arbeiten an dem Produkt/der Anlage dürfen nur im</i>

¹⁰ En este caso la suma de varios errores hacen que no se transmita el sentido del original y por ello lo marcamos como un error de tipo pragmático.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Total	Ejemplo
		<p><i>Stillstand durchgeführt werden.</i></p> <p>PE]: Los trabajos en el producto / la planta pueden ser efectuados sólo durante la parada.</p>
Adición de palabras (sin distinción del tipo de palabra añadida)	956	<p>[TA]: Ventanas siguientes de diálogo ése es señalado:</p> <p>[Original]: <i>Das folgende Dialogfenster wird angezeigt:</i></p> <p>PE]:Se indica la ventana de diálogo siguiente:</p>
Traducción inapropiada del sustantivo (contexto)	729	<p>[TA]: Usted como sistema administrador de acceso al sistema y por eso usted cuenta con todos los permisos</p> <p>[Original]: <i>Sie haben als Systemadministrator Zugang zu dem System erhalten und aus diesem Grunde verfügen Sie über alle Berechtigungen</i></p> <p>PE]:Como administrador de sistemas ha obtenido acceso al sistema y por este motivo dispone de todas las autorizaciones</p>
Traducción inapropiada del verbo (contexto)	554	<p>[TA]: La aplicación a fin de cuentas notifica, que la Deinstallation fue clausurado.</p> <p>[Original]: <i>Schließlich meldet die Anwendung, dass die Deinstallation abgeschlossen worden ist.</i></p> <p>PE]: Por último, la aplicación indica que la desinstalación ha concluido.</p>
Tratamiento erróneo de etiquetas	523	<p>[TA]: Precipitaciones incluido presiones suceda ut /.</p> <p>[Original]: Eingang Niederdruck<ut/></p> <p>[PE]: Entrada presión baja<ut/></p>
Traducción inapropiada de la preposición (contexto)	491	<p>[TA]: Puesta en servicio sólo a través de personal cualificado.</p> <p>[Original]: <i>Inbetriebnahme nur durch qualifiziertes Fachpersonal!</i></p> <p>[PE]: ¡Puesta en marcha sólo por personal especializado cualificado!</p>

Tabla 114: Errores generales más frecuentes encontrados en las traducciones

Marcamos un total de 643 errores de carácter formal. Bajo esta denominación general incluimos todos los errores que tienen que ver con la adaptación de carácter no estrictamente lingüístico como los fallos en la traducción de cifras o signos, de unidades de medida o rutas, nomenclaturas y/o etiquetas que marcan formatos como cursiva o negrita, cambio de línea, etc. El problema de las etiquetas se presenta frecuentemente ya que no se identificaron como

7. Análisis preliminar del corpus y descripción de los motores de traducción

tales y fueron tratadas como texto. Los errores pueden ser de diferente tipo, por ejemplo, pueden darse cambios en los elementos mencionados, posicionamiento erróneo, omisión o adición innecesaria de los mismos. En la tabla a continuación damos algunos ejemplos de los errores marcados. No indicamos cuántos errores se marcaron frente a cada ejemplo, porque aquí no hicimos una distinción en las categorías para el conteo, todos los errores se contaron como “errores formales” y solo los diferenciamos aquí para hacer más claro a qué nos referimos.

Tipo de error	Ejemplos
Cambios en cifras	<p>[TA]: El aparato no debe temperaturas fuera de las esferas de hasta 10 ° C + 50 ° C.</p> <p>[Original]: <i>Das Gerät darf keinen Temperaturen außerhalb der Bereiche -10 °C bis +50 °C ausgesetzt werden.</i></p> <p>[PE]: El aparato no puede ser expuesto a temperaturas fuera de los ámbitos - 10 °C hasta +50 °C .</p>
Traducción de elementos no traducibles (como URLs)	<p>[TA]: Según TÜV-Zertifikat del año demanda de energía 06.08.2009, puedo entender las www.wilo.com / legalmente.</p> <p>[Original]: <i>Jahresenergiebedarf laut TÜV-Zertifikat vom 06.08.2009, einsehbar unter www.wilo.com/legal.</i></p> <p>[PE]: Demanda energética del año según certificado de TÜV del 06.08.2009, se puede consultar en www.wilo.com/legal.</p>
Posición incorrecta de cifras	<p>[TA]: ... Si usted estos pasos 7 respetaran totalmente...</p> <p>[Original]: <i>Wenn Sie diese 7 Schritte vollständig ausgeführt haben,</i></p> <p>[PE]: Si ha completado estos siete pasos, debe reiniciar el ordenador para que los cambios tengan efecto.</p>
Tratamiento erróneo de etiquetas	<p>[TA]: Suceda ut / me en lo que se refiere a presiones por parte ...</p> <p>[Original]: <i><ut/>Bei einem Druck auf Durchsuchen können ...</i></p> <p>[PE]: <ut/>Demanda energética del año según certificado de TÜV del 06.08.2009, se puede consultar en www.wilo.com/legal.</p>

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Ejemplos
Traducción incorrecta de rutas	<p>[TA]: Por la Deinstallation de este producto tiene que suceda una función de Windows. / ut ut suceda. / retirar o modificar los programas impliquemos suceda. / ut ut / inicial (...) utilizar un sistema de gestión de programas informáticos. / ut.</p> <p>[Original]: <i>Für die Deinstallation dieses Produkts müssen Sie die Windows-Funktion <ut/> <ut/>Programme ändern oder entfernen <ut/> <ut/>(Start &gt; Systemsteuerung &gt; Software) verwenden.<ut/>.</i></p> <p>[PE]: Para desinstalar este producto debe usted emplear la función de Windows <ut/> <ut/>Cambiar o eliminar programas <ut/> <ut/> (Inicio &gt; panel de control &gt; software<ut/>.</p>

Tabla 115: Errores de carácter formal (no lingüístico)

Unos de los errores más frecuentes son las omisiones de elementos necesarios para obtener un texto natural en la lengua de llegada. Es decir, cuando en la traducción falta por completo la traducción de un elemento del original o cuando se omite una palabra —por ejemplo, una preposición o artículo— que hará clara y correcta la traducción, así dicho elemento no se encuentre en el segmento original. Clasificamos las omisiones según los tipos de palabras que no se incluyeron en el original.

Tipo de omisión	Total	Ejemplos
Ausencia de artículo	113	<p>[TA]: Imprimen gráfico en (...) formato UNE-EN 12830</p> <p>[Original]: <i>Grafik in Format UNE-EN 12830 ausdrucken</i></p> <p>[PE]: Imprimir gráfico en el formato UNE-EN 12830</p>
Ausencia de preposición	91	<p>[TA]: ... 23 millones de kwh menos (de) consumo de energía en el corredor techos bombas en Europa...</p> <p>[Original]: <i>23 Milliarden kWh weniger Stromverbrauch bei Nassläuferpumpen in Europa.</i></p> <p>[PE]: 23.000 millones de kwh menos de consumo de energía en las bombas de rotor húmedo en Europa</p>

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de omisión	Total	Ejemplos
Ausencia de verbo	7	[TA]: La persistencia de la actualización de Internet y automática a la del 0. [Original]: <i>Zum Anhalten der automatischen Aktualisierung klicken Sie auf die Ziffer 0..</i> [PE]: Para la parada de la actualización automática haga clic en la cifra 0.
Ausencia de pronombre	4	[TA]: ... permiten el control del compresor mediante (...) recogida de gas... [Original]: <i>... ermöglichen die Regelung des Kompressors durch das Aufnehmen des Gases...</i> [PE]: permiten el control del compresor mediante la recogida de gas

Tabla 116: Omisiones

Otro tipo de error frecuente es la adición de palabras que no se encuentran en el original. Las adiciones más comunes son las de palabras sin carga semántica como artículos, preposiciones y pronombres, pero también encontramos propuestas de traducción en las que se añaden sustantivos, verbos e incluso oraciones completas que no estaban en el original. Algunos motores no tratan correctamente las etiquetas como tales y al tratar de traducirlas, añaden palabras que no se encontraban en el original, lo que ocasionó que el número de errores marcados sea tan alto y por ello —aunque la causa del error la tratamos en otra categoría— lo mencionamos aquí también.

Tipo de adición	Total	Ejemplos
Adición de palabras	956	[TA]: * En Comparación con dineros prestados de calefacción desordenados [Original]: * <i>Im Vergleich zu unregulierten Heizungspumpen</i> [PE]: * En comparación con bombas de calefacción no reguladas
Adición de artículo	147	[TA]: La activación de los paquetes Adapt función [Original]: <i>Aktivierung der Dynamic Adapt Funktion</i> [PE]: Activación de la función Dynamic Adapt
Adición de pronombre	80	[TA]: Aquellos es señalado con la instalación del producto PHP 5.2.0 pregunta siguiente: [Original]: <i>Bei der Installation des Produkts <ut/> <ut/>PHP 5.2.0 <ut/> <ut/>wird die folgende Frage angezeigt:</i>

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de adición	Total	Ejemplos
		[PE]: En la instalación del producto <ut/> <ut/>PHP 5.2.0 <ut/> <ut/>se indica la cuestión siguiente:
Adición de preposición	69	[TA]: Con nuevas normas nuevas oportunidades / ErP-Direktive - 11 / 2011 [Original]: <i>Neue Chancen mit neuen Regeln / ErP-Direktive - 11/2011</i> [PE]: Nuevas oportunidades con nuevas normas / Directiva EuP - 11 / 2011
Adición de adjetivo	16	[TA]: Y son incluso Schadensersatzpflichtig, si no el estado actual de la técnica o errónea . [Original]: <i>Und sind sogar Schadensersatzpflichtig, wenn sie nicht den aktuellen Stand der Technik verbauen oder verplanen.</i> [PE]: Y tienen incluso obligación de indemnización por daños y perjuicios si no construyen o planifican según el estado actual de la técnica.
Adición de adverbio	7	[TA]: Instalación de la bomba después de clausura después del lavaje dado el caso necesario del sistema de tubo y solamente de todas Lötarbeiten y Schweiß-. [Original]: <i>Einbau der Pumpe erst nach Abschluss aller Schweiß- und Lötarbeiten und nach der gegebenenfalls erforderlichen Spülung des Rohrsystems.</i> [PE]: Antes de instalar la bomba, asegúrese de que se han completado los trabajos de soldadura y, en caso necesario, la limpieza del sistema de tuberías.
Adición de conjunción	5	[TA]: La persistencia de la actualización de Internet y automática a la del 0. [Original]: <i>Zum Anhalten der automatischen Aktualisierung klicken Sie auf die Ziffer 0,</i> [PE]: Para la parada de la actualización automática haga clic en la cifra 0.
Adición de oración	3	[TA]: Orientado a instalaciones en las cuales se requiera controlar el compresor y la válvula solenoide de líquido de forma independiente, ¹¹ lo que nos aporta un control total de la instalación (temporizaciones, protecciones, etc.). [Original]: <i>Hierdurch kann die Anlage vollständig gesteuert werden (Timer-Einstellungen, Schutzeinrichtungen, etc.).</i> [PE]: De este modo se consigue un control total de la instalación (temporizaciones, dispositivos de protección, etc.).

¹¹ En este ejemplo se añadió toda la frase marcada en rojo, que no estaba en el original.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de adición	Total	Ejemplos
Adición de verbo	2	[TA]: "mejor en clase turista"efecto global basada en un nuevo diseño solucionará Wilo-seca [Original]: " <i>Best in Class</i> " <i>Gesamtwirkungsgrade basierend auf einem neuen Wilo-Trockenläuferdesign</i> [PE]: "Best in Class" Rendimientos completos sobre la base de un nuevo diseño de rotores secos de Wilo

Tabla 117: Adiciones

En la primera fase del análisis contamos los errores teniendo en cuenta la descripción del error y el tipo de palabra que se tradujo erróneamente, p.ej. si no se traducía correctamente un sustantivo, un adjetivo o un verbo y colocamos entre paréntesis el tipo de palabra como una subcategoría para especificar el error. Así señalamos si la traducción errónea, por ejemplo, de un sustantivo se daba porque no había concordancia de género o número o porque el significado era incorrecto en el contexto o no se ajustaba a la terminología. También especificamos si el sustantivo alemán se traducía erróneamente como un adjetivo o verbo en castellano (categoría). Adicionalmente encontramos errores en la traducción de sustantivos compuestos: en algunos casos no se tradujeron todos los elementos o se alteró en la traducción el orden de éstos. Encontramos igualmente algunos casos en que la traducción era errónea porque se repetía el error del original y en algunos casos porque se daban errores ortotipográficos o por el uso incorrecto de mayúsculas y/o minúsculas. De esta manera las anotaciones realizadas también nos permitieron detallar los tipos de errores que observamos según categoría gramatical, como veremos en el apartado a continuación.

7.3.2 Tipos de errores según categorías gramaticales

En este apartado describimos los errores en función de la categoría gramatical en los que aparecían. Esta distinción es importante para nuestra tesis, porque a partir de ella podemos comprobar si las clasificaciones de errores que listamos en el marco teórico (véase página 189) incluyen todos los errores que se dan en un caso real. En la tabla a continuación listaremos los tipos de errores encontrados, los definiremos y daremos un ejemplo representativo de cada uno de ellos. En algunos casos se trata de errores en la elección de la categoría gramatical. En otros a errores de diferente naturaleza sobre una categoría gramatical determinada:

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Descripción y ejemplo
Estructura morfosintáctica de palabras, sintagmas u oración	<p>La traducción del sustantivo alemán no se ajusta a la estructura morfosintáctica del castellano, por ejemplo cuando no se traduce el caso en alemán con la preposición adecuada. En el ejemplo a continuación podemos ver como el sustantivo no se traduce correctamente porque no se reconoce el caso (genitivo) y se traduce como plural.</p> <p>[TA]: es preciso que usted una imagen telón de fondo, Internet y a la pictografía del fotografías y parece la noticia.</p> <p>[Original]: <i>Zuerst einmal müssen Sie ein Hintergrundbild anbringen, dazu klicken Sie auf das Piktogramm des Fotos und es erscheint die Nachricht:</i></p> <p>[PE]: Primero tiene que establecer una imagen de fondo, para eso haga clic sobre el pictograma de la foto y aparece el mensaje:</p>
Categoría gramatical	<p>Se traduce como otro tipo de palabra, p.ej. como verbo o adjetivo.</p> <p>[TA]: Wilo-planungshilfen proporcionan buenos argumentos para invertir en una nueva bomba</p> <p>[Original]: <i>Wilo-Planungshilfen liefern gute Argumente für Investitionen in eine neue Pumpe</i></p> <p>[PE]: Las ayudas de planificación de Wilo proporcionan argumentos buenos a favor de inversiones en una bomba nueva</p>
Adecuación al contexto	<p>La traducción es correcta pero no es la que corresponde para el contexto.</p> <p>[TA]: Prudencia</p> <p>[Original]: Vorsicht</p> <p>[PE]: Precaución</p>
Convenciones, registro o grado de formalidad	<p>La traducción es correcta pero no es la que corresponde a las convenciones culturales o del tipo de texto. Si la traducción del sustantivo alemán es para España, la traducción debe corregirse.</p> <p>[TA]: el árbol de registro de su computadora</p> <p>[Original]: <i>wird der Verzeichnisbaum Ihres Computers geöffnet</i></p> <p>[PE]: se abre el árbol de directorios de su ordenador</p>

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Descripción y ejemplo
Adecuación al tipo de texto	<p>El sentido es correcto pero la expresión no se ajusta, por ejemplo, al tipo de texto o al registro.</p> <p>[TA]: La encendida de la bomba sucede después de capítulo “entrada en servicio”</p> <p>[Original]: <i>Das Einschalten der Pumpe erfolgt nach Kapitel “Inbetriebnahme”.</i></p> <p>[PE]: Ponga en marcha la bomba según lo indicado en el capítulo “Puesta en marcha”.</p>
Flexión de género	<p>No se ajusta al género correspondiente en castellano.</p> <p>[TA]: Módulo estropeada</p> <p>[Original]: <i>Modul defekt</i></p> <p>[PE]: Módulo defectuoso</p>
Utilización de una palabra coloquial	<p>En un caso se propone una palabra que no existe en castellano:</p> <p>[TA]: En Este Caso pueden elegir el modelo respectivo después de la clicada en el aparato desconocido..</p> <p>[Original]: <i>In diesem Fall können Sie nach dem Klicken auf das unbekannte Gerät das jeweilige Modell auswählen.</i></p> <p>[PE]: En este caso, después de hacer clic en el equipo desconocido, usted puede escoger el modelo correspondiente.</p>
Flexión de número	<p>En algunos sustantivos alemanes el singular y el plural son iguales, por ello se traducen erróneamente al castellano.</p> <p>[TA]: Bombas.</p> <p>[Original]: <i>Pumpe</i>.</p> <p>[PE]: bomba.</p>
Errores provocados por una interpretación errónea de elementos lingüísticos de los segmentos originales	<p>Debido al error/problema en el original el sustantivo no se traduce correctamente.</p> <p>[TA]: Interfaces opcionales a Buskommunikation por IF-módulos einsteckbare</p> <p>[Original]: <i>Optionale Schnittstellen zur Buskommunikation durch einsteckbare IF-Module</i></p> <p>[PE]: Interfaces opcionales para comunicación de bus por medio de módulos de IF enchufables</p>

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Descripción y ejemplo
Falta de equivalencia	<p>La traducción pierde la equivalencia.</p> <p>[TA]: * frente a dineros prestados desordenados convencionales **</p> <p>[Original]: * gegenüber herkömmlichen unregelten Pumpen **</p> <p>[PE]: * en comparación con bombas convencionales no reguladas **</p>
Terminología	<p>La traducción mantiene el sentido pero no se ajusta a la terminología.</p> <p>[TA]: Potencial de ahorrar en corredores mojados en Europa:</p> <p>[Original]: <i>Einsparpotenzial bei Nassläufern in Europa:</i></p> <p>[PE]: Potencial de ahorro en rotores húmedos en Europa:</p>
No se traducen todos los elementos	<p>En el caso de sustantivos compuestos no se traducen todos los componentes.</p> <p>[TA]: dos puntos de información konfigurierbare impliquemos de explotación y Störmeldungen.</p> <p>[Original]: Zwei konfigurierbare Melderelais für Betriebs- und Störmeldungen.</p> <p>[PE]: Dos relés de señalización configurables para mensajes de servicio y averías.</p>
Orden de los elementos	<p>En el caso de sustantivos compuestos se traducen todos los componentes o se traduce alguno de ellos erróneamente.</p> <p>[TA]: El reciente anuncio de acogida y del total del consumo de electricidad.</p> <p>[Original]: <i>Anzeige der aktuellen Leistungsaufnahme und des Gesamtstromverbrauch.</i></p> <p>[PE]: Anuncio de la potencia absorbida actual y del consumo de electricidad completo.</p>

Tabla 118: Errores en la traducción de sustantivos

En la traducción de adjetivos nos encontramos además de los errores generales como puntuación o errores de convenciones con otros de errores propios de este tipo de palabras, como la flexión de número o de género:

Tipo de error	Descripción
Número	<p>No se interpreta correctamente a qué sustantivo acompaña y no se ajusta la concordancia de número correctamente.</p> <p>[TA]: Systemvordruck permisibles dentro del ámbito de aumentar.</p> <p>[Original]: <i>Systemvordruck innerhalb des zulässigen Bereiches erhöhen.</i></p>

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Descripción
	[PE]: Aumente la presión previa del sistema dentro del rango permitido .
Flexión de género	No se hace concordar al género del sustantivo en castellano. [TA]: La primero ventana es señalado con una noticia que en señala, [Original]: <i>Das erste Fenster wird mit einer Nachricht angezeigt,</i> [PE]: La primera ventana muestra un mensaje que indica,
Adecuación al contexto	La traducción es correcta pero no es la que corresponde para el contexto. [TA]: Corriente de motor demasiado alta . [Original]: <i>Zu hoher Motorstrom.</i> [PE]: La corriente del motor es demasiado elevada .
Grados del adjetivo: positivo, comparativo, superlativo	La traducción del adjetivo es correcta pero no se reproduce el grado. [TA]: El cambio en bombas de viva eficiencia se amortiza rápido que cada otra medida energética. [Original]: <i>Der Wechsel auf Hocheffizienzpumpen amortisiert sich schneller als jede andere energetische Maßnahme.</i> [PE]: El cambio a bombas de alta eficiencia se amortiza más rápido que cualquier otra medida energética.
Categoría gramatical	Se traduce por ejemplo como un adverbio o un verbo. [TA]: Fuera de eso indicado en este período de tiempo alcanzó Höchsttemperatur. [Original]: <i>Zudem wird die in diesem Zeitraum erreichte Höchsttemperatur angegeben.</i> [PE]: Además, se indica la temperatura máxima alcanzada en este período de tiempo.
No se traducen correctamente los elementos	En algunos adjetivos compuestos no se traducen correctamente algunos de los componentes. [TA]: Wilo-lcc-check ofrece la rentabilidad, anexo específicos y concretos convincentes consideraciones. [Original]: <i>Wilo-LCC-Check bietet überzeugende, anlagespezifische und konkrete Wirtschaftlichkeitsbetrachtungen.</i> [PE]: Wilo-LCC-Check ofrece análisis de rentabilidad convincentes, específicos para cada instalación y concretos.

Tabla 119: Tipos de errores en la traducción de adjetivos

7. Análisis preliminar del corpus y descripción de los motores de traducción

En la traducción de los verbos y participios encontramos errores especialmente en la traducción y conjugación de verbos separables que se componen de una raíz y un prefijo que se separan. En la mayoría de los casos el prefijo del verbo fue traducido como una preposición y no se relacionó con el verbo raíz. En los casos que no se identificó el verbo separable como tal la traducción fue incorrecta, ya que el significado del verbo depende del prefijo correspondiente y no solo de la traducción correcta de la raíz. También se presentan errores en la traducción de formas compuestas como el perfecto, el futuro o en la voz pasiva y en la traducción de oraciones en las que se combinan dos o más verbos, por ejemplo en oraciones condicionales.

Tipo de error	Descripción
Elección incorrecta del equivalente	<p>La traducción del verbo es errónea.</p> <p>[TA]: Por hacer clic en el símbolo se publican cuatro tarjetas de registro:</p> <p>[Original]: <i>Durch Klicken auf das Symbol erscheinen vier Registerkarten:</i></p> <p>[PE]: Al hacer clic en el símbolo aparecen cuatro fichas de registro:</p>
Categoría gramatical	<p>Se traduce como otro tipo de palabra, por ejemplo como sustantivo. Esto sucede cuando el verbo en alemán se escribe igual que el sustantivo y solo se reconoce en el contexto o por el uso de mayúsculas y minúsculas, como en el caso del verbo imprimir (<i>ausdrucken</i>) y el sustantivo femenino plural <i>Ausdrucken</i> (copias)</p> <p>[TA]: Gráficos en formato UNE-EN 12830 copias.</p> <p>[Original]: <i>Grafik in Format UNE-EN 12830 ausdrucken.</i></p> <p>[PE]: Imprimir gráfico en el formato UNE-EN 12830.</p>
Adecuación al contexto	<p>La traducción inapropiada para el contexto.</p> <p>[TA]: Todas las funciones se dejan cesar con el botón de servir rojo, motivan o desactivan.</p> <p>[Original]: <i>Alle Funktionen lassen sich mit dem roten Bedienknopf einstellen, aktivieren oder deaktivieren.</i></p> <p>[PE]: Todas las funciones se pueden ajustar, activar o desactivar con el botón de mando rojo.</p>
Adecuación al tipo de texto	<p>La traducción no se ajusta al tono o estilo del texto.</p> <p>[TA]: Merece la pena alta eficiencia.</p> <p>[Original]: <i>Hocheffizienz lohnt sich.</i></p>

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Descripción
	[PE]: La eficiencia alta vale la pena .
Falta de equivalencia	<p>Error en el verbo provocado por la mala interpretación de los sustantivos de la oración.</p> <p>[TA]: La posición (estado) la bomba de extracción es contadores y el consumo de electricidad en pantalla a cero.</p> <p>[Original]: <i>Die Werkseinstellung (Auslieferungszustand) der Pumpe ist wieder hergestellt und der Stromverbrauchszähler im Display steht auf Null.</i></p> <p>[PE]: La configuración de fábrica (estado del suministro) de la bomba está reestablecida y el contador de consumo de electricidad en el visualizador está en cero.</p>
Errores provocados por la redacción o las características de los segmentos originales	<p>No se traduce porque hay un error en segmento de salida.</p> <p>[TA]: De esta manera warden Usted informado de los incidentes que están con su sistema en contacto por acceso de distancia.</p> <p>[Original]: <i>Auf diese Weise warden Sie per Fernzugriff über die mit ihrem System in Verbindung stehenden Vorfälle informiert.</i></p> <p>[PE]: De este modo será usted informado por acceso remoto sobre los incidentes que guardan relación con su sistema.</p>
Terminología	<p>La traducción mantiene el sentido correcto pero no se ajusta a la terminología.</p> <p>[TA]: En este caso, usted puede detener el servidor y volver a poner en marcha,</p> <p>[Original]: <i>In diesem Falle können Sie den Server stoppen und neu starten,</i></p> <p>[PE]: En este caso puede parar y reiniciar el servidor ..</p>

Tabla 120: Tipos de errores en la traducción de los verbos

Los errores en la traducción de los verbos se producen también por errores en la conjugación. Clasificamos los errores en la conjugación de acuerdo al punto en que se daban. Es decir, si no se flexionaban en el modo, tiempo, voz o persona correcta. Casi todos los motores tuvieron problemas con la identificación del modo, ya que en alemán, por ejemplo, el infinitivo, la primera y la tercera persona plural de los verbos regulares en el modo indicativo suelen ser idénticas y también muchos participios perfectos no se diferencian del infinitivo.

Encontramos errores de traducción de artículos, pronombres, preposiciones, conjunciones y adverbios que se presentan en categorías similares a por ejemplo las de los adjetivos.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Descripción
Categoría	Se traduce como otro tipo de palabra, por ejemplo ciertas preposiciones en alemán pueden funcionar también como adverbios y solo se pueden traducir correctamente teniendo en cuenta el contexto.
Contexto	La traducción es correcta pero no es la que corresponde para el contexto
Significado	Cuando la traducción es errónea.
Estructura morfosintáctica de palabras, sintagmas u oración	No se ajusta la traducción a la estructura morfosintáctica del castellano, por ejemplo cuando no se traduce el caso en alemán con la preposición adecuada.
Convenciones	La traducción es correcta pero no es la que corresponde a las convenciones culturales o del tipo de texto
Flexión de género	No se ajusta al correspondiente en castellano o se traduce incorrectamente desde el comienzo por homonimia como en el caso el artículo masculino singular <i>der</i> que se escribe igual que el femenino en el caso genitivo.
Número	No se ajusta al correspondiente en castellano o se traduce incorrectamente desde el comienzo por homonimia
Tipo	Cuando se traduce al castellano la misma categoría de palabra pero un tipo diferente. Por ejemplo se utiliza un pronombre demostrativo, mientras que en el original se emplea un pronombre indefinido.
Posición	Si la palabra traducida se coloca en la posición gramaticalmente incorrecta. En los casos puntuales indicamos entre paréntesis qué tipo de palabra se colocó erróneamente en la oración.

Tabla 121: Errores en la traducción de otros tipos de palabras

Sin tener en cuenta el motor con que se realizaron las traducciones, hicimos un listado de los errores más frecuentes. En la primera fase del análisis como lo indicamos anteriormente tuvimos en cuenta todas las propuestas de traducción y no solo las que se escogieron para la posesición (véase anexo en el CD-ROM: Primera fase.xls).

7. Análisis preliminar del corpus y descripción de los motores de traducción



Ilustración 38: Porcentaje de errores más frecuentes independientemente del motor de traducción empleado

En la gráfica podemos ver que la mayoría de los errores en este análisis fueron aquellos en que no había coherencia y esto impedía la comprensión del texto. En la tabla a continuación daremos ejemplos de acuerdo al tipo y también indicaremos el total de errores encontrados. No solo aparecerá un ejemplo ilustrando la traducción automática, sino también el segmento original y su posesición.

Tipo de error	Total	Ejemplo
Orden sintáctico erróneo de los elementos en la oración	1954	[TA]: Usted éste cliquean aparece indicación de pantalla siguiente: [Original]: <i>Sie diesen anklicken erscheint folgende Bildschirmanzeige:</i> [PE]: Al hacer clic en éste aparece la siguiente indicación en pantalla:
Error global de sentido (impide la comprensión del mensaje)	1794	[TA]: Los arrancadores de rapidez están más bien en el objetivo. [Original]: <i>Schnellstarter sind eher am Ziel.</i> [PE]: Los que toman la delantera alcanzan antes el objetivo.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Total	Ejemplo
No se traduce correctamente el equivalente (terminología)	1165	[TA]: Comprobar protecciones . [Original]: <i>Sicherungen überprüfen.</i> [PE]: Compruebe los fusibles .
Adición de palabras	956	[TA]: Obtienen aquellos Bildschirmanzeige siguiente: [Original]: <i>Sie erhalten die folgende Bildschirmanzeige:</i> [PE]: Recibirá la siguiente indicación en pantalla:
Falta de traducción del sustantivo	945	[TA]: No pueden hacerlo a través de los correspondientes Pfeile hacia arriba o hacia abajo. [Original]: <i>Sie können es über die entsprechenden Pfeile nach oben oder unten bewegen.</i> [PE]: Puede moverlo mediante las flechas correspondientes hacia arriba o hacia abajo.

Tabla 122: Errores más frecuentes detectados en el primer análisis

Una vez que hemos descrito las principales categorías de errores abordaremos cuáles fueron los errores que se encontrarán más frecuentemente dependiendo del tipo de motor utilizado para la traducción de los originales del corpus.

7.4 Errores encontrados según el motor empleado

Ya que nuestro interés está centrado en los errores que se corrigen en la posesición de traducciones automáticas no describiremos, aunque los marcamos, los errores que encontramos las propuestas de la memoria de traducción puesto que no son —según nuestro conocimiento— traducciones automáticas. Tampoco nos referiremos a los errores que no se corrigieron en la posesición.

7.4.1 Errores encontrados en las propuestas del motor SMT1

Con los siguientes ejemplos queremos ilustrar cuáles fueron los errores más frecuentes en las propuestas de traducción realizadas con este motor. Marcamos un total de 5090 errores. El problema principal que detectamos es que las oraciones no tienen sentido y les falta coherencia (570 errores), entre otras por los otros errores que se presentan, por ejemplo porque se añaden palabras (395 errores) o no se traducen nombres o sustantivos porque se desconoce la terminología (293 errores). También encontramos propuestas en las que el orden sintáctico de

7. Análisis preliminar del corpus y descripción de los motores de traducción

los elementos en la oración no es correcto (450 errores) y casos en los que no se tradujo el sustantivo (248 errores). Debemos anotar que se añadieron palabras generalmente en los casos en que las etiquetas no se reconocieron como tales y se trataron de traducir (253 errores).



Ilustración 39: Errores más frecuentes en las propuestas de SMT1

En la siguiente tabla daremos algunos ejemplos de los tipos de errores que encontramos más frecuentemente. Como en las tablas anteriores también indicaremos la cantidad de errores marcados:

Tipo de error	Total	Ejemplo
Error global de sentido (impide la comprensión del mensaje)	570	[TA]: Entre [Original]: <i>Bedienung</i> [PE]: Manejo
Orden sintáctico erróneo de los elementos en la oración	450	[TA]: Si apagamos hidráulicos en los flujos muy pequeño volumen la bomba en un al respecto. [Original]: <i>Bei sehr kleinen Volumenströmen geht die Pumpe hierzu in einen hydraulischen Standby.</i> [PE]: Con flujos volumétricos muy pequeños la bomba se pone en modo de espera hidráulico.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Total	Ejemplo
Adición de palabras	395	[TA]: y configuración sobre un "botón rojo" Seis-Cuatro y pantalla [Original]: <i>Bedienung und Konfiguration über bekannte "Roter Knopf"-Technologie und Display</i> [PE]: Mando y configuración mediante la tecnología de "botón rojo" conocida y visualizador
Traducción inapropiada del sustantivo (terminología)	293	[TA]: A una red ferroviaria todas las contrataciones y anuncios en el de almacenamiento. [Original]: <i>Bei einer Netzunterbrechung bleiben alle Einstellungen und Anzeigen im Speicher erhalten.</i> [PE]: En caso de una interrupción de red todos los ajustes y anuncios permanecen en la memoria.
Tratamiento erróneo de etiquetas	253	[TA]: " las cosas field name = " \$texto " / me " en la página de hacer las cosas field name = "\$pagenum" / me [Original]: " <i><field name="\$paratext"/> " auf Seite <field name="\$pagenum"/></i> " [PE]: " <i><field name="\$paratext"/></i> " en la página <i><field name="\$pagenum"/></i>
Falta de traducción del sustantivo	248	[TA]: En Wärmedämmarbeiten vigilar que el motor y el módulo no esté- [Original]: <i>Bei Wärmedämmarbeiten darauf achten, dass der Pumpenmotor sowie das Modul nicht gedämmt werden.</i> [PE]: Durante la realización de tareas de aislamiento térmico , asegúrese de no aislar el motor de la bomba ni el módulo.

Tabla 123: Ejemplos de los errores encontrados en la SMT1

7.4.2 Errores encontrados en las propuestas del motor SMT2

Marcamos un total de 5072 errores en las propuestas realizadas por este motor. Podemos ver que los errores que más se presentan son casi los mismos que encontramos en las propuestas realizadas por el motor smt1. El problema principal también es que las traducciones no tienen sentido y les falta coherencia (551 errores), entre otras por los otros errores que se presentan, por ejemplo porque se añaden palabras (394 errores) o no se traducen nombres o sustantivos porque se desconoce la terminología (323 errores). También encontramos propuestas en las que el orden sintáctico de los elementos en la oración no es correcto (509 erro-

7. Análisis preliminar del corpus y descripción de los motores de traducción

res); casos en que las etiquetas no se reconocieron como tales y se trataron de traducir (242 errores) y otros casos en los que no se tradujo el sustantivo (233 errores). En la ilustración a continuación podemos ver los seis tipos de errores más frecuentes y en qué porcentaje con respecto al total de errores marcados se presentaron.

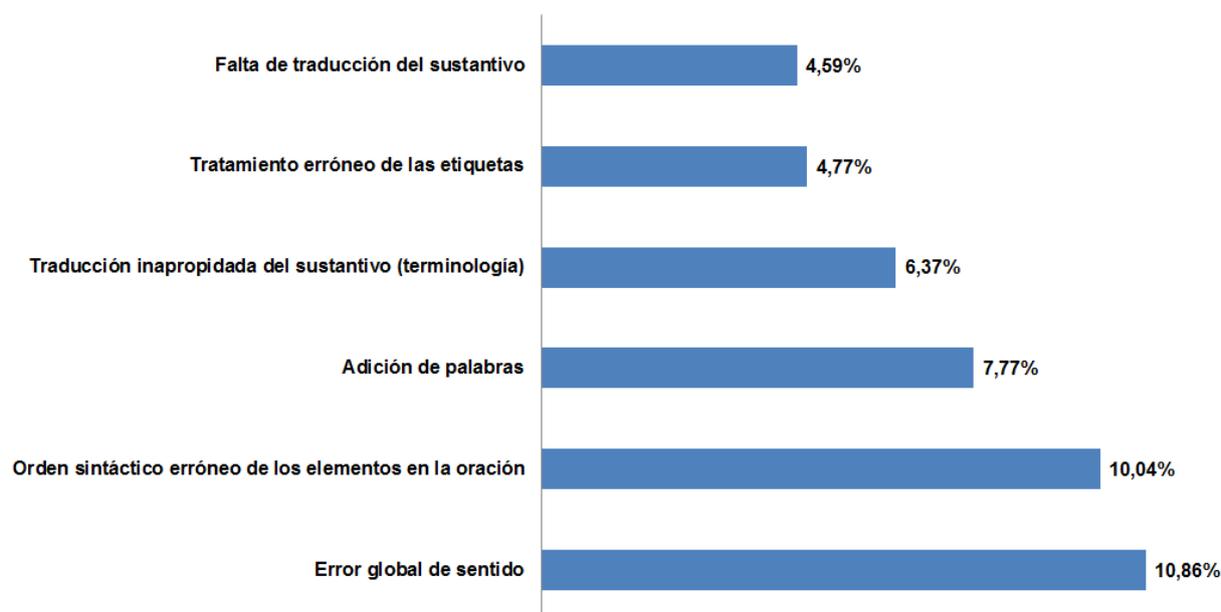


Ilustración 40: Errores más frecuentes en las propuestas de SMT2

A continuación daremos algunos ejemplos para ilustrar los errores que encontramos, también indicaremos la cantidad de errores marcados.

Tipo de error	Total	Ejemplo
Error global de sentido (impide la comprensión del mensaje)	551	[TA]: ... el nivel de financiación en. [Original]: <i>Einstellung der Förderhöhe in m.</i> [PE]: Ajuste de la altura de impulsión en m.
Orden sintáctico erróneo de los elementos en la oración	509	[TA]: impliquemos ahora instalado el sistema sucesivamente los diferentes productos que en AKO 5004 [Original]: <i>Ab jetzt installiert das System nacheinander verschiedene Produkte, die an AKO 5004 beteiligt sind.</i> [PE]: A partir de ahora el sistema instala uno tras otro diferentes productos que intervienen en AKO 5004
Adición de palabras	394	[TA]: ... Sistema símbolo de riesgos.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Total	Ejemplo
		[Original]: <i>Allgemeines Gefahrensymbol</i> [PE]: Símbolo de peligro general
Traducción inapropiada del sustantivo (terminología)	323	[TA]: Reducción de la noche [Original]: <i>Nachtabsenkung</i> [PE]: Reducción nocturna
Tratamiento erróneo de las etiquetas	242	[TA]: Suceda ut /: Respeto. [Original]: <i><ut/>ACHTUNG:</i> [PE]: <i><ut/> ATENCIÓN:</i>
No se traduce el sustantivo	233	[TA]: Lo hará usted una Klappliste con todos los interfaces de comunicación su Servers... [Original]: <i>Es wird Ihnen eine Klappliste mit allen verfügbaren Kommunikationsschnittstellen ihres Servers angezeigt....</i> [PE]: Se le indica una lista plegable con todas las interfaces de comunicación disponibles de su servidor...

Tabla 124: Ejemplos de los errores encontrados en las propuestas de SMT2

7.4.3 Errores encontrados en las propuestas del motor RBMT1

A diferencia de las traducciones de los sistemas estadísticos las propuestas de este motor son comprensibles aunque en algunos casos tengan problemas de precisión por errores de terminología o porque las traducciones de ciertas palabras no se ajustan al contexto. Hallamos además problemas de orden sintáctico, pero estos también se dan con menor frecuencia que en los sistemas basados en estadísticas. Como indicamos en la página 131 las propuestas de este motor fueron las que presentaron menos errores, en total marcamos 3081 faltas. En la ilustración vemos los problemas más frecuentes y el porcentaje en qué los encontramos con respecto al total de errores marcados.

7. Análisis preliminar del corpus y descripción de los motores de traducción

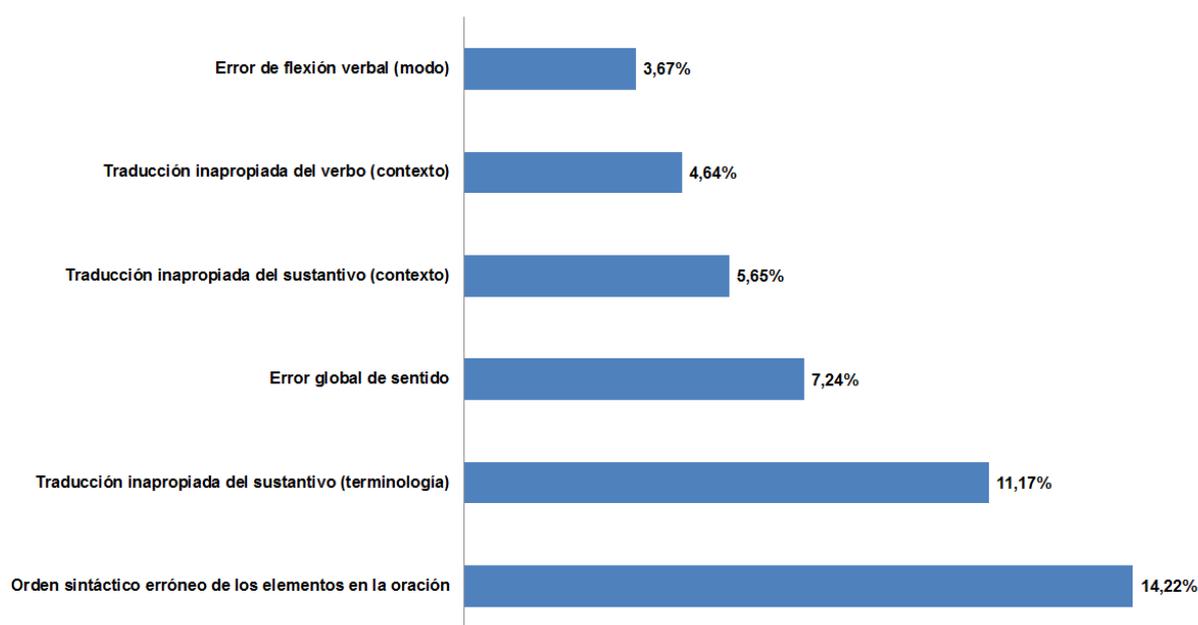


Ilustración 41: Errores más frecuentes en las propuestas del motor RBMT1

A continuación algunos ejemplos representativos de los errores más frecuentes e indicamos el número de errores marcados.

Tipo de error	Total	Ejemplo
Error de orden sintáctico	438	[TA]: En caso de utilización del rebajamiento nocturno tiene que estar instalada en el inicio del sistema de calefacción la bomba. [Original]: <i>Bei Nutzung der Nachtabsenkung muss die Pumpe im Vorlauf des Heizungssystems installiert sein.</i> [PE]: En caso de utilización de la reducción nocturna, la bomba tiene que estar instalada en la fase inicial del sistema de calefacción.
Traducción inapropiada del sustantivo (terminología)	344	[TA]: Se le indica una lista de funcionar con todas las interfaces de comunicación disponibles ... [Original]: <i>Es wird Ihnen eine Klappliste mit allen verfügbaren ...</i> [PE]: Se le indica una lista plegable con todas las interfaces de comunicación disponibles ...

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Total	Ejemplo
Error global de sentido	223	[Original]: * <i>Im Vergleich zu unregulierten Heizungspumpen</i> [TA]: * En Comparación con dineros prestados de calefacción desordenados [PE]: * En comparación con bombas de calefacción no reguladas
Traducción inapropiada del sustantivo (contexto)	174	[TA]: Componentes que mueven defensa de contacto para sí (p. ej. acoplamiento) puede ser alejado en sí en la empresa producto existente no. [Original]: <i>Berührungsschutz für sich bewegende Komponenten (z.B. Kupplung) darf bei sich im Betrieb befindlichem Produkt nicht entfernt werden.</i> [PE]: La protección contra contacto accidental de los componentes móviles (p. ej., el acoplamiento) no debe ser retirada del producto mientras éste se encuentre en funcionamiento .
Traducción inapropiada del verbo (contexto)	143	[TA]: En Wilo vivido fácil. [Original]: <i>Bei Wilo gewohnt einfach.</i> [PE]: Con Wilo es fácil.
Error de flexión verbal (modo)	113	[TA]: Detener el servidor AKO5004 y los cierran usted. [Original]: <i>Halten Sie den Server AKO5004 an und schließen Sie ihn.</i> [PE]: Pare el servidor AKO5004 y ciérrelo .

Tabla 125: Ejemplos de los errores encontrados en las propuestas de RBMT1

7.4.4 Errores que se encuentran en las propuestas del motor RBMT2

Aquí nos encontramos de nuevo con traducciones sin sentido y con problemas de orden gramatical. Asimismo, no se traducen todas las palabras o se traducen de manera que no se ajustan al contexto, lo que ocasiona en algunos casos que el sentido sea confuso.

7. Análisis preliminar del corpus y descripción de los motores de traducción

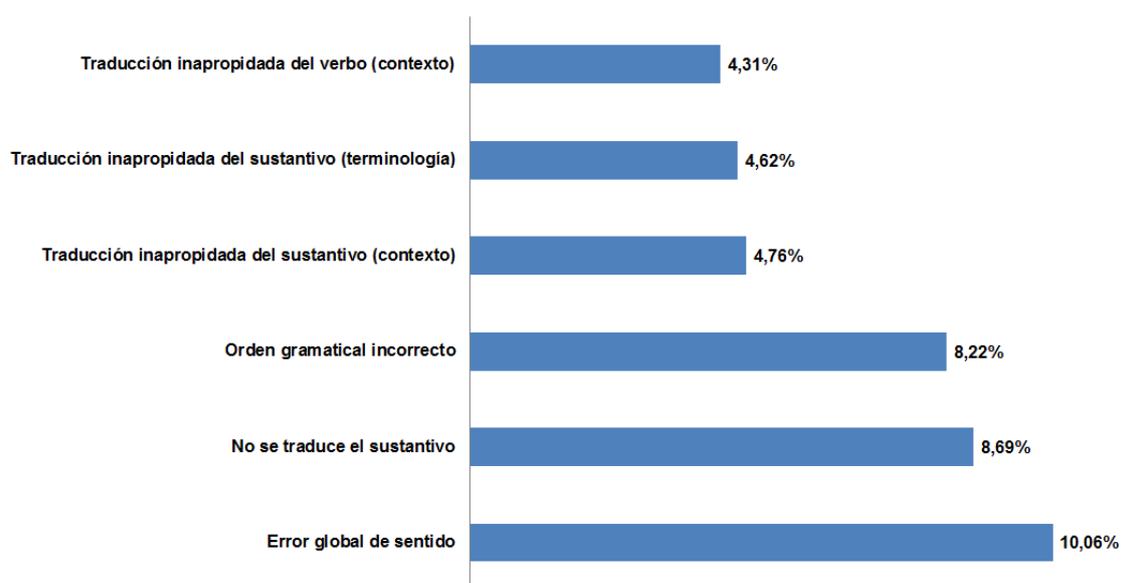


Ilustración 42: Errores más frecuentes en las propuestas de RBMT2

Como podemos ver en la gráfica en las propuestas de este sistema se presentan muchos errores provocados por la falta de equivalente en el sustantivo, ya sea porque no se traduce el equivalente indicado para el contexto o porque no se respeta la terminología, sino también porque incluso se dejan de traducir. Este tipo de errores deberían poderse subsanar mediante una gestión del léxico especializado específico del texto original, de modo que no supongan ningún problema para el poseedor. A continuación presentamos algunos ejemplos de los errores encontrados más frecuentemente:

Tipo de error	Total	Ejemplo
Orden gramatical incorrecto	557	[TA]: Además ventana siguiente es señalado: [Original]: <i>Dazu wird folgendes Fenster angezeigt:</i> [PE]: Para ello se indica la ventana siguiente:
Error global de sentido	450	[TA]: El programa de ons de ati de establo de tu señala una fila de ventanas con informaciones de este momento a la Deinstallation del producto. [Original]: <i>Das Deinstallationsprogramm zeigt ab diesem Moment eine Reihe von Fenstern mit Informationen zur Deinstallation des Produkts an.</i> [PE]: El programa de desinstalación enseña a partir de este momento una serie de ventanas con informaciones acerca de la desinstalación del producto.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Tipo de error	Total	Ejemplo
Falta de traducción del sustantivo	385	<p>[TA]: El aparato mismo conduce, lee el valor Unterdruckwächters también así como, y el compresor así como el electroimán de líquido.</p> <p>[Original]: Das Gerät selbst steuert sowohl den Kompressor als auch den Flüssigkeitselektromagneten, und liest auch den Wert des Unterdruckwächters ab.</p> <p>[PE]: El equipo se encarga de controlar tanto el compresor como el solenoide de líquido, así como de leer el estado del presostato de baja presión.</p>
Traducción inapropiada del sustantivo (contexto)	213	<p>Ahora mismo intercambian impermeabilizaciones defectuosas.</p> <p>[Original]: Defekte Dichtungen sofort austauschen.</p> <p>PE. Si detecta daños en alguna junta, cámbiela inmediatamente.</p>
Traducción inapropiada del sustantivo (terminología)	205	<p>[TA]: La bomba (Fig. 1/1) consistís de una hidráulica, un motor de corredor de mojado con rotor de imán de permanente y Regelmodul electrónico con Frequenzumrichter integradas.</p> <p>[Original]: Die Pumpe (Fig. 1/1) besteht aus einer Hydraulik, einem Nassläufermotor mit Permanentmagnetrotor und einem elektronischen Regelmodul mit integriertem Frequenzumrichter.</p> <p>[PE]: La bomba (Fig. 1/1) está compuesta por un sistema hidráulico, un motor de rotor húmedo con rotor de imán permanente</p>
Traducción inapropiada del verbo (contexto)	191	<p>[TA]: Éste posibilita un pilotaje más preciso bombea down, puesto que Verzögerungszeiten y la seguridad pueden ser contratado con esta atención.</p> <p>[Original]: Dies ermöglicht eine präzisere Steuerung der pump down, da die Sicherheits und Verzögerungszeiten bei dieser Bedienung eingestellt werden können.</p> <p>[PE]: Esto permite un control más preciso de la recogida de gas, ya que nos permite definir los tiempos de seguridad y retardo de esta maniobra.</p>

Tabla 126: Ejemplos de errores encontrados en las propuestas de RBMT2

7. Análisis preliminar del corpus y descripción de los motores de traducción

7.5 Comparación de los sistemas

La tabla de conteo de errores también nos permite ver las categorías de los errores que más fueron marcados y cómo se comportó cada sistema. De esta manera pudimos listar los diez errores más frecuentes y la manera en que se presentaban según el motor. Así pudimos ver que casi solo los sistemas estadísticos tuvieron problemas con las etiquetas, que interpretaron como texto y generaron errores de traducción. También se puede ver que el motor RBMT2 dejó más sustantivos sin traducir que los otros motores y que los motores estadísticos presentan más errores de coherencia que los motores basados en reglas. Otra observación interesante es que a pesar de ser el motor más favorecido por los poseedores, las propuestas del RBMT 1 presentan más problemas de fluidez que los otros motores. A continuación la tabla detallada del número de errores, en la que marcaremos por categoría el motor que más errores presentó.

Tipo de error	smt1	smt2	rbmt1	rbmt2
Orden sintáctico erróneo de los elementos en la oración	450	509	438	557
Error global de sentido	570	551	223	450
Traducción inapropiada del sustantivo (terminología)	295	324	346	207
Falta de traducción del sustantivo	263	235	79	389
Adición de palabras	395	394	80	87
Traducción inapropiada del sustantivo (contexto)	168	174	175	213
Traducción inapropiada del verbo (contexto)	112	103	146	193
Tratamiento erróneo de las etiquetas	253	242	7	21
Traducción inapropiada de la preposición (contexto)	125	136	113	123
Flexión verbal incorrecta (modo)	102	101	118	167

Tabla 127: Comparación de los errores más frecuentes y los motores

Según el número de errores las propuestas menos favorecidas son las del rbmt2 del SMT1; en las propuestas de estos motores encontramos más errores. Mientras que los errores que presentan las propuestas del motor RBMT1 tienen más que ver con el orden sintáctico y la terminología. Estos resultados se pueden ver en la siguiente ilustración:

7. Análisis preliminar del corpus y descripción de los motores de traducción

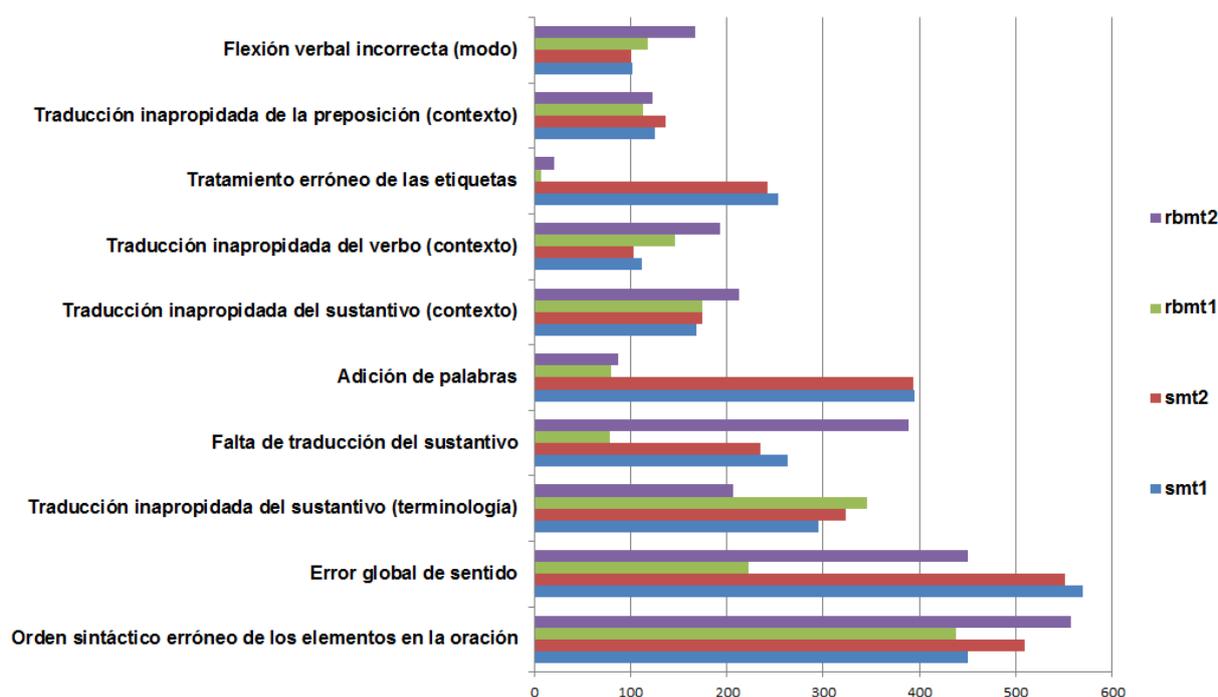


Tabla 128: Comparación de los errores más frecuentes y los motores empleados

7.6 Análisis de las traducciones con ayuda de la distancia de teclado (*Levenshtein*)

Como lo indicamos en el apartado “Reorganización de los datos del corpus” (página 220) gracias a los datos obtenidos pudimos elaborar una tabla en la que contamos el número de palabras de cada propuesta de traducción y así pudimos calcular la distancia Levenshtein al texto poseído. Este análisis nos reveló detalles interesantes. En primer lugar observamos que en 15 de los 940 segmentos traducidos se optó por traducir de cero y no se poseitó ninguna de las propuestas de los motores de traducción automática. Como enunciamos al comienzo del apartado en la página 217, no todos los segmentos analizados tenían una propuesta de la memoria de traducción; en estos casos encontramos siempre una copia del original en alemán frente a la etiqueta “tm”. Así pudimos reconocer los segmentos que fueron traducidos y no poseídos, porque aparecía marcada como propuesta seleccionada para la posesición la salida de “tm”, que en estos casos era la copia del original. En segundo lugar, pudimos observar que en 121 casos se optó por una propuesta de un motor, sin razón especial, a pesar de que otro y en algunos casos varios motores habían arrojado la misma propuesta que se utilizó para la posesición. En tercer lugar, encontramos 105 casos en los que el poseedor no optó por corregir la traducción con menos cambios (menor distancia Levenshtein) sino que

7. Análisis preliminar del corpus y descripción de los motores de traducción

optó por poseer otra propuesta de traducción. Es decir, el poseedor no elige automáticamente la propuesta que le obliga a hacer menos ediciones. Veamos unos ejemplos:

Sistema	Original o propuesta de traducción	LV
Source	<F id="23">* gegenüber herkömmlichen unregelmäßigen Pumpen ** gegenüber herkömmlichen regulierten Pumpen </F>	
[TM]:	<F id="23">* gegenüber herkömmlichen unregelmäßigen Pumpen ** gegenüber herkömmlichen regulierten Pumpen </F>	87
[smt1]:	Las cosas F id = "12" de hacer las cosas Tab / me me suceda Tab / me ahorro potencial de Nassläufern en Europa: las cosas / F me	99
[smt2]:	Las cosas F id = "23" me * frente a las bombas irregulares tradicionales * * frente a las bombas regulado tradicionales de hacer las cosas / F me	88
[rbmt1]:	<F id="23">* frente a bombas convencionales desorganizadas frente a bombas convencionales reguladas ** </F>	73
[rbmt2]:	<F id="23">* frente a bombas convencionales desorganizadas frente a bombas convencionales reguladas ** </F>	46
postedited	<F id="23">* en comparación con bombas convencionales no reguladas ** en comparación con bombas convencionales reguladas </F>	

Tabla 129: Ejemplos en los que el poseedor prefirió traducir que poseer

En este ejemplo podemos ver que para el poseedor era más fácil traducir la oración que corregir las salidas que tienen muchos problemas. A continuación daremos un ejemplo de uno de los casos en que no encontramos una razón específica para entender por qué el poseedor seleccionó una propuesta específica si había otras que eran idénticas. En algunos de los 121 casos pudimos ver que en los casos en que una de las opciones era una propuesta de la memoria de traducción, el poseedor optó por seleccionar para la posesión esta propuesta. Pero también tuvimos otros casos en los que no se podría determinar un criterio exacto. Como en la tabla de análisis de datos, marcamos la propuesta seleccionada para la posesión en color azul:

Sistema	Original o propuesta de traducción	LV
Source	Automatische Deinstallation des Produkts	
[TM]:	Automatische Deinstallation des Produkts	27

7. Análisis preliminar del corpus y descripción de los motores de traducción

Sistema	Original o propuesta de traducción	LV
[smt1]:	Deinstallation automática del producto	4
[smt2]:	Deinstallation automática del producto	4
[rbmt1]:	Deinstallation automática del producto	4
[rbmt2]:	Deinstallation automáticas del producto	5
postedited	Desinstalación automática del producto	

Tabla 130: Ejemplo de selección de una propuesta en particular teniendo otras opciones iguales

Aquí podemos ver que las propuestas de los motores SMT1, SMT 2 y RBMT1 son idénticas, pero el poseedor selecciona la primera, tal vez por ese motivo, porque le apareció en primer lugar. Finalmente daremos un ejemplo de uno de los casos en que el poseedor selecciona una propuesta que tiene una distancia Levenshtein mayor:

Sistema	Original o propuesta de traducción	LV
Source	<i>Klicken Sie auf die Schaltfläche <ut/>Next<ut/>:</i>	
[TM]:	<i>Klicken Sie auf die Schaltfläche <ut/>Next<ut/>:</i>	27
[smt1]:	Clic usted la jugar las cosas ut / me Next suceda ut / me:	41
[smt2]:	Hagan un clic en los botones impliquemos impliquemos Next. / ut ut /:.	43
[rbmt1]:	Haga sobre la superficie de conectar clic <ut/>Next<ut/>:	29
[rbmt2]:	Clicar en la Schaltfläche<ut/> Next<ut/>:	22
postedited	Haga clic en el botón <ut/>Next<ut/>:	

Tabla 131: Ejemplo de selección de una propuesta con una distancia mayor según Levenshtein

En el ejemplo podemos ver que las propuestas del motor RBMT1 y el RBMT 2 tienen un valor menor de una distancia, sin embargo el poseedor corrige la propuesta del SMT2. No se puede determinar por qué el poseedor no opto por una de las propuestas con una distancia menor. Pero a pesar de tratarse de que el resultado del test indique que la distancia en el caso del SMT2 es mayor, la estrategia del poseedor pudo consistir en eliminar el texto sobrante a partir de la parte correcta del inicio del segmento, por lo que la distancia de tecleo en este caso podrían no refleja con exactitud las acciones emprendidas por el traductor. Es por eso que el valor de la distancia de teclado no nos puede ayudar a identificar en todos los casos cuáles son los segmentos que el poseedor prefiere.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Sin embargo en la ilustración a continuación podemos ver en porcentajes que las respuestas con una menor distancia fueron preferidas por los poseedores.

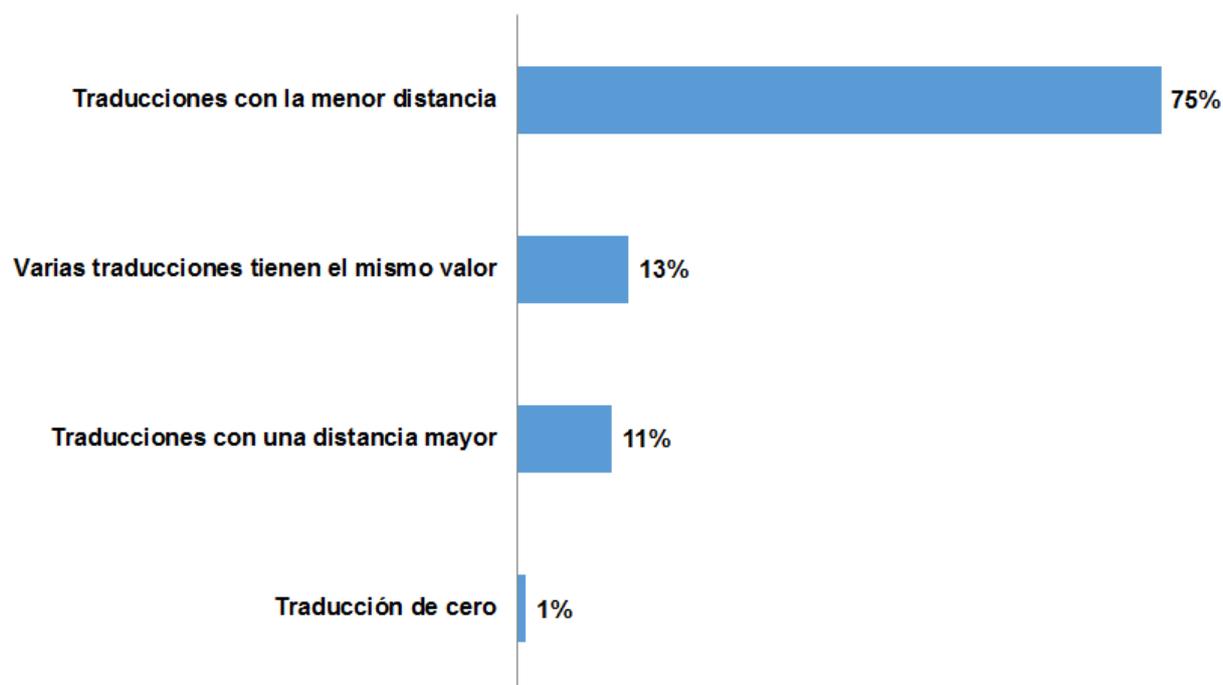


Ilustración 43: Porcentaje de traducciones según el valor de la distancia de teclado

7.7 Errores presentes en el corpus según las categorías de MQM

Para tener una perspectiva adicional sobre la categorización de errores hicimos, como lo indicamos anteriormente, una revisión del corpus utilizando las categorías de errores de MQM (ver apartado Anexo en el CD-ROM). En los apartados a continuación explicaremos detalladamente los resultados de dicha clasificación.

7.7.1 Errores encontrados en la categoría Accuracy

La categoría *Accuracy* se especifica en siete categorías principales, que denominamos en nuestra tabla categorías de primer rango, y dos de ellas tienen subcategorías y por ello las que denominamos categorías de segundo rango:

Categoría primer rango	Categoría segundo rango		Total
<i>Addition</i>			191
<i>Improper-exact-tm-match</i>			0
<i>Mistranslation</i>	<i>(general)</i>	154	

7. Análisis preliminar del corpus y descripción de los motores de traducción

Categoría primer rango	Categoría segundo rango		Total
	<i>Ambiguous-translation</i>	0	
	<i>Date/time</i>	0	
	<i>Entity</i>	14	
	<i>False friend</i>	0	
	<i>Mistranslation of technical relationship</i>	288	
	<i>Number</i>	0	
	<i>Overly literal</i>	36	
	<i>Should not have been translated</i>	14	
	<i>Unit conversion</i>	0	
			490
<i>Omission</i>			143
	<i>Omitted variable</i>	0	
<i>Over-translation</i>			0
<i>Under-translation</i>			1
<i>Untranslated</i>			123

Tabla 132: Número de errores encontrados en la categoría *Accuracy*

Como podemos ver la mayoría de los errores marcados se encuentran en la categoría *Mistranslation* (490). Un error de la categoría *Mistranslation* se da, según MQM, cuando el contenido de destino no representa con precisión el contenido original. Es por ello que cuando nos encontramos con una traducción errónea en general la marcamos como un error de esta categoría, sin especificar ninguna subcategoría. En segundo lugar se encuentran las adiciones y en tercer lugar las omisiones de palabras o contenido. También marcamos en esta categoría las palabras que no se tradujeron (*Untranslated*).

7. Análisis preliminar del corpus y descripción de los motores de traducción

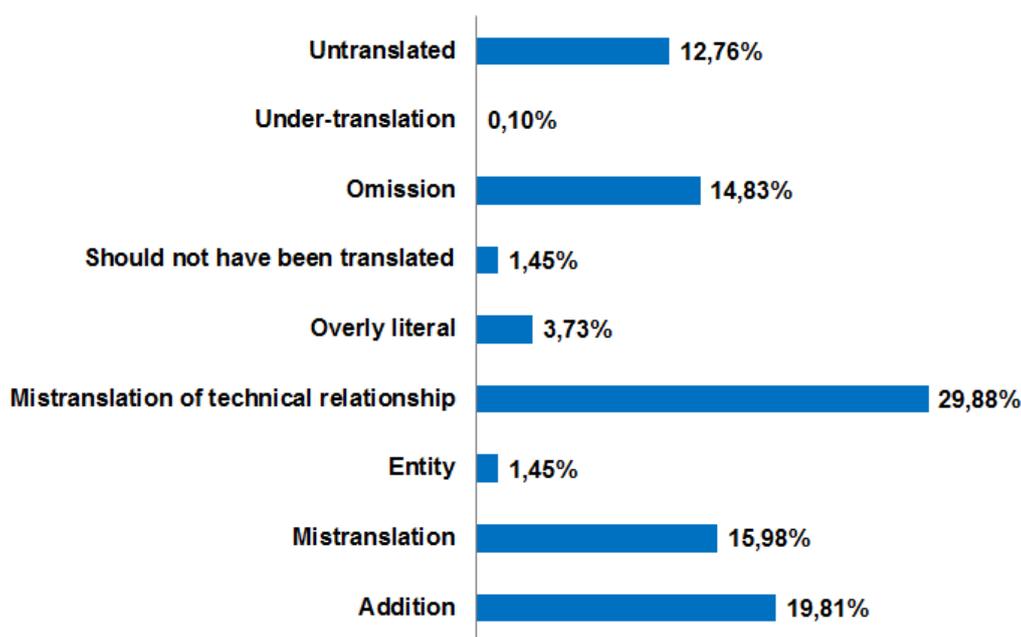


Ilustración 44: Errores más frecuentes encontrados en la categoría *Accuracy*

En la tabla a continuación daremos ejemplos representativos de los errores marcados en esta categoría:

Categoría de error	Número	Ejemplo
<i>Mistranslation (general)</i>	490	[TA]: Atrasar en supresión de obra . [Original]: Auf <i>Werkseinstellung</i> zurücksetzen. [PE]: Volver a la configuración de fábrica .
<i>Addition</i>	190	[TA]: Los datos sobre el producto. [Original]: <i>Angaben über das Erzeugnis</i> . [PE]: Datos sobre el producto
<i>Omission</i>	147	[TA]: Con un la conexión existe riesgo mediante la descarga eléctrica. [Original]: Bei <i>unsachgemäßem elektrischem Anschluss besteht Lebensgefahr durch Stromschlag</i> . [PE]: Con una conexión inadecuada existe peligro de muerte debido a electrocución..
<i>Untranslated</i>	121	[TA]: Deinstallation automática del producto. [Original]: Automatische <i>Deinstallation</i> des Produkts. [PE]: Desinstalación automática del producto

Tabla 133: Ejemplos de errores encontrados en la categoría *Accuracy*

7. Análisis preliminar del corpus y descripción de los motores de traducción

Bajo la categoría *Mistranslation* nos encontramos con que la mayoría de los errores se marcaron en la subcategoría *mistranslation of technical relationship*. Ello se debe a que con esta categoría marcamos los errores que se presentaron cuando no se encontró el equivalente adecuado, porque se trataba de un término técnico. Algunos de estos términos son compuestos nominales que no aparecen en el diccionario, otros son palabras polisémicas que tienen un significado específico en el contexto técnico. También marcamos en esta subcategoría los errores que se dieron en la traducción de nombres propios (*Entity*), las palabras que se tradujeron innecesariamente (*Should not have been translated*), las traducciones que fueron excesivamente literales y no se adaptan al contexto o sintaxis del idioma meta (*Overly literal*).

En la figura siguiente ilustramos gráficamente la distribución proporcional de los errores encontrados en esta subcategoría.

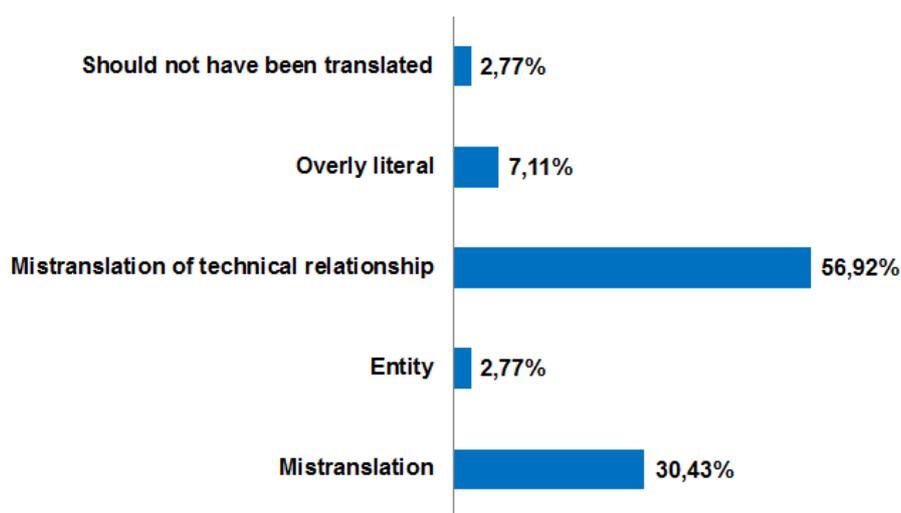


Ilustración 45: Errores encontrados en la subcategoría *Mistranslation*

A continuación daremos ejemplos representativos de los errores que marcamos bajo esta subcategoría:

Categoría de error	Número	Ejemplo
<i>Mistranslation of technical relationship</i>	274	[TA]: Después de que haya seleccionado el idioma deseado, hace sobre la superficie de conectar clic. [Original]: Nachdem Sie die gewünschte Sprache ausgewählt haben, klicken Sie auf die Schaltfläche .

7. Análisis preliminar del corpus y descripción de los motores de traducción

Categoría de error	Número	Ejemplo
		[PE]: Después de que haya seleccionado el idioma deseado, haga clic sobre el botón
<i>Mistranslation (general)</i>	152	[TA]: Un 20% del consumo de electricidad universal de motores corresponde a Pumpen [Original]: 20% des weltweiten Stromverbrauchs von Motoren entfallen auf Pumpen. [PE]: Un 20% del consumo mundial de electricidad de motores corresponde a las bombas.
<i>Overly literal</i>	38	[TA]: Apostar ahora ya por Productos Alto Eficientes. [Original]: Jetzt schon auf Hocheffiziente Produkte setzen.. [PE]: Apostar ahora ya por productos altamente eficientes.
<i>Should not have been translated</i>	13	[TA]: Haga clic a continuación sobre la superficie de conectar <ut/>Acabado<ut/>. [Original]: Klicken Sie anschließend auf die Schaltfläche <ut/>Finish<ut/>. [PE]: Haga clic a continuación sobre el botón <ut/>Finish<ut/>.
<i>Entity</i>	13	[TA]: anual promedio del consumo de electricidad sólo 46,5 kwh del sur de ITV certificados [Original]: Durchschnittlicher Jahresstromverbrauch von nur 46,5 kWh vom TÜV SÜD zertifiziert [PE]: Promedio del consumo anual de electricidad sólo 46,5 kwh certificado por TÜV SÜD * *

Tabla 134: Ejemplos de los errores encontrados en la subcategoría *Mistranslation*

7.7.2 Errores encontrados en la categoría *Design*

La segunda categoría del MQM Core es la categoría *Design*; aquí marcamos solo errores en la subcategoría *Markup*, que son los errores relacionados con las etiquetas. Las otras categorías y subcategorías no son relevantes para nuestro corpus ya que tienen que ver con aspectos no lingüísticos de los cuales no tenemos ninguna información como maquetación, color o tipo de letra, etc. La categoría *Markup* tiene cuatro subcategorías que denominamos en la tabla general como categorías de segundo rango. Estas categorías son *Added-markup*, que se da cuando se ha añadido una etiqueta sin margen de beneficio correspondiente; las subca-

7. Análisis preliminar del corpus y descripción de los motores de traducción

tegorías *Misplaced-markup* y *Missing-markup* se explican por su nombre mismo. También encontramos la subcategoría *questionable-markup* que se utiliza para marcar los errores que se presentan cuando la etiqueta está presente, pero aparece en un formato incorrecto o inadecuado para su contexto. Bajo esta categoría marcamos los errores que se dieron en el corpus cuando se trataron las etiquetas como texto y por ello se dañaron.

Categoría primer rango	Categoría segundo rango		Total
<i>Markup</i>			39
	<i>Added-markup</i>	0	
	<i>Misplaced-markup</i>	1	
	<i>Missing-markup</i>	2	
	<i>Questionable-markup</i>	36	

Tabla 135: Número de errores encontrados en la categoría *Design*

En la tabla a continuación daremos un par de ejemplos de los errores marcados:

Categoría de error	Número	Ejemplo
<i>Questionable markup</i>	38	[TA]: Para que la aplicación funcione correctamente, es necesario un relanzamiento del ordenador. ut / me. [Original]: <i>Damit die Anwendung ordnungsgemäß funktioniert, ist ein Neustart des Computers erforderlich.</i> <ut/> [PE]: Para que la aplicación funcione correctamente es necesario reiniciar el ordenador. <ut/>
<i>Misplaced markup</i>	1	[TA]: <ut/>Modifican<ut/><ut/> escalas de tamaño. [Original]: <ut/>Größenskalen ändern<ut/> <ut/>. [PE]: <ut/>Modificar escalas de tamaño<ut/> <ut/>.
<i>Missing markup</i>	2	[TA]: Suceda Frm ID = ""17"". / suceda. B: Reducción de la noche suceda / b. [Original]: <Frm id="17"/> Nachtabsenkung:. [PE]: <Frm id="17"/> Reducción nocturna:.

Tabla 136: Ejemplos de errores encontrados en la categoría *Design*

7. Análisis preliminar del corpus y descripción de los motores de traducción

7.7.3 Errores encontrados en la categoría *Fluency*

Bajo esta categoría clasifican la tipología de MQM los errores relacionados con la forma o el contenido de un texto, independientemente de si se trata de una traducción o no. En esta categoría tiene 18 subcategorías (categorías de primer rango) y algunas de ellas tienen otras subcategorías (categorías de tercer rango). Aquí marcamos principalmente errores que tenían que ver con la gramática, puntuación y ortografía. En la tabla a continuación listaremos solo las subcategorías de *Fluency* en las que marcamos errores, con el número de errores encontrados.

Categoría primer rango	Categoría segundo rango	Categoría tercer rango		Total
<i>Duplication</i>				1
<i>Grammar</i>				709
	<i>Function words</i>		189	
	<i>Word-form</i>		313	
		<i>Agreement</i>	111	
		<i>Part of speech</i>	33	
		<i>Tense-mood-aspect</i>	169	
	<i>Word-order</i>		207	
<i>Inconsistency</i>				
	<i>Inconsistent abbreviations</i>		6	6
<i>Spelling</i>				53
	<i>Capitalization</i>		52	
	<i>Diacritics</i>		1	
<i>Typography</i>			8	71
	<i>Punctuation</i>		49	
	<i>Unpaired-marks</i>		4	
	<i>Whitespace</i>		10	
<i>Unintelligible</i>				21

Tabla 137: Número de errores encontrados en la categoría *Fluency*

7. Análisis preliminar del corpus y descripción de los motores de traducción

Como podemos ver en la tabla la mayoría de los errores encontrados se presentaron en las categorías *Grammar*, *Typography* y *Spelling*.

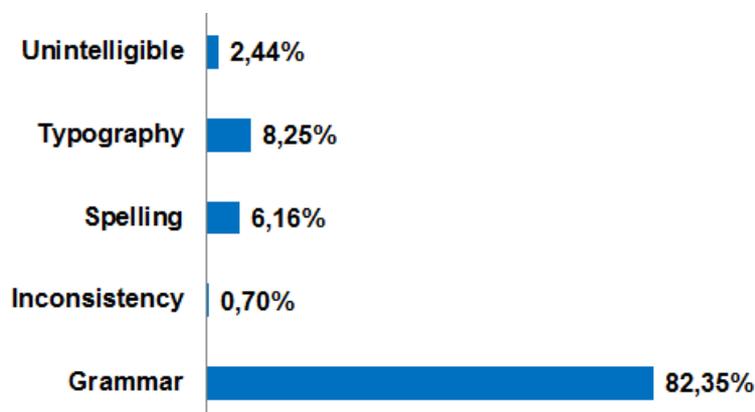


Ilustración 46: Errores encontrados en la categoría *Fluency*

Podemos hacer una distinción en las subcategorías de errores de MQM, lo que es especialmente importante para la subcategoría *Grammar*, que se divide en las categorías de segundo rango *Function words*, *Word-form* y *Word-order*. Bajo la subcategoría *Function words* marcamos los errores que se dieron cuando se utilizaron de forma incorrecta preposiciones, verbos auxiliares, artículos y determinantes. La categoría *Word form* agrupa los errores que tiene que ver con la forma de las palabras y se divide en tres subcategorías o categorías de tercer rango en nuestra tabla. En primer lugar encontramos la subcategoría *Agreement*, que se refiere a los errores que se dan cuando dos o más palabras no concuerdan en caso, número, persona u otra función gramatical. En segundo lugar tenemos la subcategoría *Part of speech* que se aplica a los casos en que una palabra no se ajusta al contexto y por último la categoría *Tense-mood-aspect* bajo la que se clasifican los errores en la flexión verbal. Finalmente encontramos la categoría de segundo orden *Word order* bajo la que se pueden clasificar los errores de sintaxis. A continuación daremos ejemplos que ilustran los errores que marcamos en dicha subcategoría:

Categoría de error	Número	Ejemplo
<i>Function words</i>	185	[TA]: De vuelta a la Wizard [Original]: <i>Zurück zum Wizard</i> [PE]: Volver al asistente.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Categoría de error	Número	Ejemplo
<i>Word form :Agreement</i>	104	[TA]: Puede añadir las variables de todos los aparatos que están conectadas en la red. [Original]: <i>Sie können die Variablen aller Geräte hinzufügen, die am Netz angeschlossen sind.</i> [PE]: Puede añadir las variables de todos los aparatos que están conectados a la red
<i>Word form:Part of speech</i>	30	[TA]: Actualizado de una tarjeta de registro [Original]: <i>Registerkarte Aktualisierung</i> [PE]: Actualización de la ficha de registro
<i>Word-form :Tense-mood-aspect</i>	151	[TA]: Señalan puntos, esconden [Original]: <i>Punkte anzeigen/ausblenden</i> [PE]: Mostrar puntos, ocultar
<i>Word-order</i>	198	[TA]: Automáticamente actualizan [Original]: <i>Automatisch aktualisieren</i> [PE]: Actualizar automáticamente.

Tabla 138: Errores de la subcategoría *Grammar*

Bajo *Fluency* también encontramos errores en la subcategoría *Spelling* y *Typography*. En *Spelling* se hace una distinción entre los errores en el uso de mayúsculas y minúsculas y en la utilización de signos diacríticos. En *Typography* se diferencia entre errores de puntuación, de uso de comillas y de espacios en blanco.

Categoría de error	Número	Ejemplo
<i>Capitalization</i>	52	[TA]: * basada en la combinación energética de la UE [Original]: * <i>Basierend auf EU-Energiemix</i> [PE]: * Basada en la combinación energética de la UE,
<i>Diacritics</i>	1	[TA]: Cómo ve, puede ser asignado en casa de este usuario nuevo ya un nombre y un login [Original]: <i>Wie Sie sehen, kann bei diesem neuen Anwender schon ein Name und ein Login zugeteilt</i> [PE]: Como ve, ya le puede asignar a este usuario nuevo un nombre y un login

7. Análisis preliminar del corpus y descripción de los motores de traducción

Categoría de error	Número	Ejemplo
<i>Punctuation</i>	55	[TA]: /los trabajos al producto, la planta pueden ser efectuado sólo en la paralización. [Original]: <i>Die Arbeiten an dem Produkt/der Anlage dürfen nur im Stillstand durchgeführt werden.</i> [PE]: Los trabajos en el producto / la planta pueden ser efectuados sólo durante la parada.
<i>Unpaired-marks</i>	5	[TA]: Servicio y configuración sobre tecnología de botón " "de rojos conocida y visualizador [Original]: <i>Bedienung und Konfiguration über bekannte "Roter Knopf"-Technologie und Display</i> [PE]: Mando y configuración mediante la tecnología de "botón rojo" conocida y visualizador
<i>Whitespace</i>	11	[TA]: La Directiva vigente desde 2009 (2009 / 125 / CE). [Original]: <i>Die seit 2009 geltende Richtlinie (2009/125/EG)</i> [PE]: La Directiva, vigente desde 2009 (2009/125/CE)

Tabla 139: Errores de las subcategorías *Spelling* y *Typography*

Por último encontramos en la categoría *Fluency* la subcategoría *Unintelligible* bajo la que marcamos todos los segmentos que no tenían sentido:

Categoría de error	Número	Ejemplo
<i>Unintelligible</i>	21	[TA]: Wilo-lcc-check ofrece la rentabilidad, anexo específicos y concretos convincentes consideraciones. [Original]: <i>Wilo-LCC-Check bietet überzeugende, anlagespezifische und konkrete Wirtschaftlichkeitsbetrachtungen.</i> [PE]: Wilo-LCC-Check ofrece análisis de rentabilidad convincentes, específicos para cada instalación y concretos

Tabla 140: Errores de la subcategoría *Unintelligible*

7.7.4 Errores encontrados en la categoría *Internationalization*

Como enunciamos anteriormente no encontramos ningún error en esta categoría.

7. Análisis preliminar del corpus y descripción de los motores de traducción

7.7.5 Errores encontrados en la categoría *Locale convention*

Aquí no encontramos sino tres errores en total, uno de los cuales tiene que ver con el formato de moneda.

Categoría de error	Número	Ejemplo
<i>Currency format</i>	1	[TA]: Gastos totales en € [Original]: <i>Gesamtkosten in EUR</i> [PE]: Gastos totales en EUR

Tabla 141: Errores de la subcategoría *Locale convention*

7.7.6 Errores encontrados en la categoría *Style*

Los errores clasificados en esta categoría de MQM se refieren a lo que comúnmente se conoce como “estilo”, que se define de manera formal (en las guías de estilo) e informalmente como un estilo claro, bueno, fluido. Estas cuestiones están estrechamente relacionadas con la fluidez, pero a menudo son tratadas por separado por las herramientas y los procesos de calidad y por ello en MQM se agrupan como una dimensión separada. En esta categoría marcamos en total 28 errores.

Categoría de error	Número	Ejemplo
<i>Style (general)</i>	5	[TA]: Se tienen que excluir peligros por energía eléctrica. [Original]: <i>Gefährdungen durch elektrische Energie sind auszuschließen.</i> [PE]: Se deben excluir los peligros debidos a la energía eléctrica.
<i>Register</i>	0	
<i>Register: Variants-slang</i>	0	
<i>Awkward</i>	9	[TA]: Con la eliminación adecuada y reciclado se este producto daños medioambientales y una amenaza para la salud personal., [Original]: <i>Mit der ordnungsgemäßen Entsorgung und des sachgerechten Recycling dieses Produktes werden Umweltschäden und eine Gefährdung der persönlichen Gesundheit vermieden.</i> [PE]: Con la eliminación adecuada y el reciclado correcto de este producto se evitarán daños medioambientales y los riesgos para la salud personal.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Categoría de error	Número	Ejemplo
<i>Company style</i>	0	
<i>Inconsistent style</i>	0	
<i>Third-party-style</i>	1	<p>[TA]: Deinstallieren usted ataque HTTP servidor 2.2.3. para la Deinstallation importante que este producto la Windows-Funktion programas o eliminar (sistema de lanzamiento me gestión me software).</p> <p>[Original]: <i>Deinstallieren Sie Apache HTTP Server 2.2.3. Für die Deinstallation dieses Produkts müssen Sie die Windows-Funktion Programme ändern oder entfernen (Start > Systemsteuerung > Software) verwenden.</i></p> <p>[PE]: Para la desinstalación de este producto debe usted emplear la función de Windows Cambiar o eliminar programas (Inicio > panel de control > software).</p>
<i>Unidiomatic</i>	8	<p>[TA]: Hagan un clic en elaboración, con el fin de mostrar gráficos.</p> <p>[Original]: <i>Klicken Sie auf Erstellen, um die Grafik anzuzeigen.</i></p> <p>[PE]: Hagan clic en “crear” con el fin de mostrar el gráfico:</p>

Tabla 142: Errores de la subcategoría *Style*

Los errores que marcamos como errores de *Third-party-style* son aquellos que se dieron cuando, por ejemplo, no se respetó el estilo de Microsoft para describir rutas de programas. Como errores en la subcategoría *Unidiomatic* marcamos aquellos casos en que el contenido es gramatical, pero no idiomático.

7.7.7 Errores encontrados en la categoría *Terminology*

Bajo esta categoría marcamos 247 errores, en sentido general y marcamos en particular el uso inconsistente de la terminología en el corpus, ya que se utilizaron varios términos para un mismo concepto en el original y también encontramos diferentes traducciones de un mismo término. También errores que se dieron por la terminología específica del texto. Especialmente la traducción errónea de compuestos nominales que no aparecen en un diccionario corriente y que solo son conocidas en un contexto especializado:

7. Análisis preliminar del corpus y descripción de los motores de traducción

Categoría de error	Número	Ejemplo
<i>Terminology (general)</i>	1	[TA]: En todo el mundo, sola bomba corredor techos regulada por el uso de aire acondicionado [Original]: <i>Weltweit einzige geregelte Nassläuferpumpe für den Einsatz in Kälte-/Klimaanlagen</i> [PE]: En todo el mundo la única bomba de rotor húmedo regulada para el uso en instalaciones refrigeración / climatización
<i>Inconsistente with term-base</i>	0	
<i>Inconsistent with domain</i>	239	[TA]: Potencial de ahorrar en corredores mojados en Europa: [Original]: <i>Einsparpotenzial bei Nassläufern in Europa:</i> [PE]: Potencial de ahorro en rotores húmedos en Europa:
<i>Inconsistent use of terminology</i>	1	[TA]: Controlar la directriz de ErP con Wilo eficientemente altamente [TA]: En la gran complejidad de la Directiva ErP le apoyamos.
<i>Multiple translations of same term</i>	15	[TA]: Los datos de los registradores se ponen cada 3 horas automáticamente abajo [TA]: Podemos actualizar la aplicación con los mejora en la tarjeta de registro actualización [TA]: Todos los parámetros en varios mapas de registro

Tabla 143: Errores de la subcategoría *Terminology*

7.7.8 Errores encontrados en la categoría *Verity*

En la categoría *Verity*, bajo la que se clasifican los errores que se presentan cuando se hace declaraciones que contradicen el mundo ‘exterior’ del texto, marcamos solo dos errores. Uno hace referencia a los requerimientos legales (*Legal requirements*), es decir cuando un texto no cumple los requisitos legales que se dan normalmente en las especificaciones y errores de la subcategoría de tercer rango de *End-user-suitability* ya que la traducción no es adecuada para su uso, porque el usuario no podría saber de qué se trata y por ejemplo no encontrar la información deseada. En nuestro ejemplo sobre la categoría *Legal requirements* por ejemplo que la norma no se traduce correctamente a la norma correspondiente en castellano. No encontramos errores en las demás subcategorías de *Verity*: *completeness*, *culture-specific references* y *local-specific content*.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Categoría de error	Número	Ejemplo
<i>End-user-suitability</i>	1	[TA]: "<field name=""\$paratext"/>" en lado <field name="\$pagenum"/> [Original]: "<field name=""\$paratext"/>" <i>auf Seite</i> <field name="\$pagenum"/> [PE]: "<field name=""\$paratext"/>" en la página <field name="\$pagenum"/>
<i>Legal-requirements</i>	1	[TA]: Nuevas oportunidades con nuevas normas / ErP-directiva - 11 / 2011. [Original]: <i>Neue Chancen mit neuen Regeln / ErP-Direktive - 11/2011</i> [PE]: Nuevas oportunidades con nuevas normas / Directiva EuP - 11 / 2011

Tabla 144: Errores de la subcategoría *Verity*

7.8 Conclusiones del primer análisis

El conteo de errores (MQM) es muy granulado y está dirigido al diagnóstico de sistemas de traducción automática y también al control de calidad de la traducción humana, además tiene en cuenta la calidad del texto origen y su métrica se basa en una evaluación de la calidad de una traducción específica. Pero dicho conteo de errores en general demanda mucho tiempo, requiere entrenamiento, especialmente para poder distinguir entre los tipos de errores, no todos los errores son igualmente relevantes para todos los proyectos.

El haber establecido una tipología tan amplia de errores nos dificulta el hacer un estudio comparativo en el que podamos ver cómo se ha comportado cada sistema teniendo en cuenta los segmentos poseditados. Sin embargo, una vez realizada la primera clasificación tuvimos una vista general de los errores que se presentaron. El haber marcado los errores en todas las traducciones y no solo en las que se escogieron para la posesición nos permitió ver en qué motor se presentaron más errores y de esta manera qué motor funcionó mejor para la traducción de este tipo de textos. Con esta clasificación general de los datos nos dimos también cuenta que una descripción detallada de los errores, si bien es útil en un primer momento para detectarlos y clasificarlos, no es práctica para la clasificación general de los mismos.

7. Análisis preliminar del corpus y descripción de los motores de traducción

Además estos resultados no son los que necesitamos para nuestro propósito, porque nuestro objetivo no es describir los errores únicamente en función del resultado, sin tener en cuenta el original, ya que esto puede provocar apreciaciones imprecisas. Nuestro foco no es formular una nueva tipología de errores sino ayudar a los poseedores en nuestra combinación lingüística y la sistematización de los problemas de cara a una toma de decisiones ordenada no facilita la toma de decisiones equiparables ante diferentes segmentos que presentan la misma categoría de error. En este sentido podemos concluir que muchos de los errores descritos no pueden abordarse únicamente desde la perspectiva de la lengua de llegada para facilitar la generación de estrategias de resolución de problemas, ya que están relacionados o bien con la lengua o la redacción del texto de partida, o bien con las diferencias entre la lengua de partida y la lengua de llegada.

8. Clasificación de errores para la sistematización de la posesición

En la primera fase de análisis del corpus (capítulo anterior) describimos cómo actuaron los motores y qué errores detectaron los poseedores en todas las propuestas de traducción. Esta segunda fase del análisis está orientada a detectar qué tipo de problemas se editan, cuál es su origen, de qué manera se editan y con qué estrategia. Ya que nuestro foco está dirigido no a la evaluación de los diferentes motores de traducción, sino al análisis de los cambios que se realizan en el proceso de la posesición, en esta fase solo analizaremos las traducciones que fueron poseídas. Es decir, abordaremos los errores de tal manera que podamos esclarecer de dónde provienen y en qué momento vale la pena intervenir, para poder establecer —siguiendo nuestro objetivo principal— normas o directrices que contribuyan en la realización del trabajo de posesición en nuestra combinación lingüística. En esta fase no tendremos en cuenta los posibles errores que haya cometido el poseedor, como tampoco los errores que pasaron inadvertidos en el proceso de posesición.

A continuación describiremos el procesamiento de datos que seguimos para este análisis. Para ello en el primer aparte describiremos la manera en que procesamos los datos poseídos a partir de las traducciones automáticas. Luego describiremos los elementos utilizados para la clasificación de los errores y los valores con que determinados cada error. Luego detallaremos de manera jerárquica los tipos de problemas que se fueron identificados, en forma de macrocategorías y categorías de errores.

8.1 Procesamiento de los segmentos poseídos a partir de TA

Para la segunda fase del análisis hicimos una copia del archivo de cálculo que utilizamos para la anotación de los de errores (véase página 213) y que ya empleamos en la primera fase. En este archivo eliminamos, en primer lugar, todos los registros en los cuales se escogió la propuesta de la memoria de traducción para la posesición y, en segundo lugar, solo dejamos para cada segmento la propuesta que fue escogida para ser poseída. Así obtuvimos una hoja de cálculo con 651 segmentos originales, las propuestas seleccionadas y la posesición de la traducción correspondiente. Para el conteo de errores creamos una hoja de cálculo con la siguiente estructura:

8. Clasificación de errores para la sistematización de la posesición

Columna	Contenido
Número	Nueva numeración de los segmentos, después de borrar aquellos en los que se escogió la propuesta de la memoria de traducción para la posesición.
Número original	Aquí aparece el número de identificación de este segmento en la hoja de cálculo inicial, en la que se encuentran todas las traducciones y todos los segmentos (incluso aquellos en los que se escogió la propuesta de la memoria de traducción para la posesición). Este número que aparece al lado de cada segmento en el archivo que utilizamos para la primera fase del análisis.
Sistema	En esta columna aparece la descripción del contenido de la columna siguiente, es decir si se trata del segmento original aparecerá <source> y si se trata de la traducción automática aparecerá el motor que se escogió para la posesición, es decir <smt1>, <smt2>, <rbmt1> o <rbmt2> y también aparece <posedited> cuando al frente, es decir en la columna siguiente, encontramos el segmento poseeditado.
Source	En esta columna aparece, según lo que se aparezca el segmento original en alemán.
Posedited	En esta columna aparece el segmento original poseeditado
MT	Aquí se encuentra la traducción automática del correspondiente segmento
Descripción	En esta columna aparece una breve descripción del error
Categoría	Aquí se encuentra el número de la macrocategoría y la subcategoría del error encontrado. Una descripción detallada de las categorías se podrá ver a partir de la página 285.
Valor del error	Aquí aparece el tipo de error. La descripción detallada se podrá leer en el apartado “Valores de los errores” a partir de la página 301.

Tabla 145: Descripción de las columnas de la hoja de cálculo que utilizamos para la segunda fase del análisis

En esta fase definimos en primer lugar seis categorías principales o macrocategorías de error, que agrupan a su vez categorías más específicas. Bajo las macrocategorías abordamos los aspectos generales en los que se puede cometer errores, aquí nos referiremos a) en la primera lugar a los errores que se dan en cualquier combinación lingüística; b) en la segunda macrocategoría agrupamos los errores relacionados con las características propias del alemán; c) en la tercera los errores relacionados con las características del castellano; d) en la cuarta macrocategoría los errores que se dan en la microestructura de los segmentos u oraciones; e) en la quinta macrocategoría agrupamos los errores que se dan por problemas o errores en el original y f) por último dejamos una macrocategoría general en la que agrupamos los errores que no se puedan clasificar en ninguna de las macrocategorías mencionadas anteriormente.

8. Clasificación de errores para la sistematización de la posesición

8.2 Descripción de los elementos de la clasificación de errores

En los apartados a continuación presentaremos nuestra clasificación concreta en macrocategorías y categorías de errores, así como los valores de errores que asignamos a cada edición que se realizó en una determinada categoría. No entraremos en detalle en cada categoría sino ofreceremos un cuadro general de estas. En el apartado 8.3 haremos una definición detallada y con ejemplos.

8.2.1 Macrocategorías y categorías de errores

Nuestra clasificación, como indicamos anteriormente, está orientada a ayudar a solventar cada error y por ello se diferencia de otro tipo de clasificaciones (véase tipologías de errores a partir de la página 131). Las clasificaciones que mencionamos en el marco teórico buscan describir los errores, sobre todo del texto traducido/revisado por humanos, mientras que la clasificación que proponemos busca establecer qué error se ha cometido en la traducción automática para a) proporcionar directrices claras al poseedor sobre cómo actuar ante una traducción en nuestra combinación lingüística y b) intentar identificar de dónde surge el error y, si es posible —siguiendo el objetivo de nuestra tesis— detectar la manera de ayudar a los poseedores el poder evitarlo. En la tabla a continuación podemos ver las categorías que se agrupan bajo cada macrocategoría:

Macrocategorías	Categorías
Primera macrocategoría: errores independientes de la combinación lingüística:	
	a) Naturaleza lingüística: <ol style="list-style-type: none">1) Presencia de léxico extranjero2) Tratamiento de nombres propios, abreviaturas, siglas
	b) Naturaleza extralingüística: <ol style="list-style-type: none">1) Errores en la conversión de convenciones como unidades de medida, números, moneda, etc.2) Errores de funcionalidad de elementos tipográficos3) Errores en las referencias (enlaces, vínculos, hipertexto)
Segunda macrocategoría: errores provocados por las características propias del alemán:	
	a) Verbos compuestos separables
	b) Conjunciones y partículas

8. Clasificación de errores para la sistematización de la posesición

c) Errores de transferencia sintáctica de composiciones

d) Errores en oraciones compuestas:

d.1) Errores en la traducción de oraciones subordinadas

d.2) Errores en la traducción de oraciones coordinadas

e) Errores en la traducción de la negación

f) Errores en la traducción de colocaciones o fraseologías

Tercera macrocategoría: errores provocados por características del castellano:

a) Errores de flexión o concordancia:

1) Género

2) Número

b) Errores de flexión verbal:

1) Tiempo

2) Modo

3) Número

4) Voz

c) Errores de registro/estilo

d) Errores ortotipográficos:

1) Errores ortográficos

2) Errores tipográficos

e) Errores de sintaxis

Cuarta macrocategoría: errores en la microestructura:

a) Palabras sin carga semántica:

1) Errores de carácter preposicional

1.1) Errores en la traducción de preposiciones

1.2) Errores en la traducción de estructuras preposicionales

2) Errores en la traducción de determinantes

3) Errores de carácter adverbial

3.1) Errores en la traducción de adverbios

3.2) Errores en la traducción de estructuras adverbiales

4) Errores en la traducción de conjunciones y partículas

8. Clasificación de errores para la sistematización de la posesición

<ul style="list-style-type: none">b) Palabras con carga semántica:<ul style="list-style-type: none">1) Errores en la traducción de sustantivos2) Errores en la traducción de adjetivos3) Errores en la traducción de verbos<ul style="list-style-type: none">3.1) Errores en la traducción de verbos simples3.2) Errores en la traducción de formas verbales compuestas4) Errores relativos a compuestos<ul style="list-style-type: none">4.1) Errores en la transferencia sintáctica de compuestos nominales4.2) Errores en la fluidez de los sintagmas5) Errores en la transferencia léxica de compuestos nominales y composiciones
Quinta macrocategoría: errores relacionados con género del texto o la redacción del texto original
<ul style="list-style-type: none">a) Presencia de etiquetas en el original
<ul style="list-style-type: none">b) Registro/estilo del texto original
<ul style="list-style-type: none">c) Error ortotipográfico en el original
Sexta macrocategoría: otros errores

Tabla 146: Macrocategorías y categorías de errores empleadas en la segunda fases del análisis

Las macrocategorías cumplen la función de organizar los problemas que encontramos y nos ayudan a clasificarlos desde un punto de vista más genérico. Esta clasificación intenta ser exhaustiva y recoger todos los posibles errores posibles en posesición de traducciones entre alemán y castellano. Sin embargo, no cabe duda de sólo hemos podido describir aquellos errores presentes en nuestro corpus de análisis, por lo que la estructura de análisis podría resultar en algún aspecto incompleta.

Después de analizar los resultados de la primera fase y de identificar las macrocategorías y categorías mencionadas, también pudimos darnos cuenta de que en cada categoría pueden darse diversos valores a los errores, que pueden variar dependiendo de la categoría misma. En otras palabras, en la segunda fase no solo clasificamos los errores según macrocategorías y categorías, sino que además tuvimos en cuenta las diferentes concreciones posibles de cada error, que hemos denominado valores y que nos ayudan a calificar las categorías mismas.

8. Clasificación de errores para la sistematización de la posesición

8.2.2 *Valores de los errores*

En la tabla a continuación listamos y explicamos cada tipo o valor que se puede dar a un error encontrado. Mientras que con las categorías definimos los problemas que se dan en la traducción, con los valores determinamos cuál es el error que se da y por qué motivo no se resuelve correctamente el problema respectivo. Como la mayoría de las tipologías de error utilizan valores similares en inglés para describir cómo se concreta cada error, hemos decidido denominar los errores de acuerdo con los términos utilizados por ejemplo en métricas como MQM (a partir de la página 154).

Valor del error	Descripción
Mistranslation	El contenido de destino no representa con precisión el contenido de origen o es completamente erróneo.
Omission	Falta algún contenido en la traducción que está presente en el texto original.
Addition	El texto de destino incluye texto que no está presente en el texto original.
Untranslated	El contenido que debería haber sido traducido se ha dejado sin traducir.
Questionable	La etiqueta o signo, por ejemplo de puntuación aparece malformado o no es adecuada o adecuado para el contexto.
Misplaced	La etiqueta o signo, por ejemplo de puntuación, se encuentra en el lugar inapropiado.
Not idiomatic	La traducción no refleja el uso concreto de la palabra o expresión correspondiente.
Register	La traducción, o el equivalente seleccionado en caso de una palabra polisémica, no es la empleada en el contexto concreto.

Tabla 147: Descripción de los tipos de errores

Habiendo descrito los dos conceptos principales que utilizamos en la segunda fase del análisis del corpus, haremos una descripción de cada macrocategoría y categoría, con ejemplos y con los valores de errores que pueden presentarse.

8.3 *Definiciones por macrocategorías y categorías de errores*

A continuación describimos detalladamente todas las categorías de errores identificados en el corpus a partir de la macrocategoría a la que pertenecen.

8. Clasificación de errores para la sistematización de la posesición

8.3.1 Primera macrocategoría: errores independientes de la combinación lingüística

Aquí se encuentran todos los errores que podrían presentarse en cualquier traducción automática independientemente de su combinación lingüística. Nos referimos, por ejemplo, a la presencia de léxico extranjero —que en la primera fase denominamos presencia de anglicismos— al tratamiento de los nombres propios o a los problemas que se presentan en la traducción o conversión de unidades de medida o errores ortotipográficos. Estos errores pueden ser de tipo lingüístico o extralingüístico.

Categoría 1.a: errores de naturaleza lingüística

Bajo esta categoría marcamos los errores que se dieron básicamente por dos problemas principales: la presencia de léxico extranjero y el tratamiento inadecuado de los nombres propios, abreviaturas y/o siglas. Entendemos que estos dos errores son de naturaleza equivalente porque en ambos casos se trata de unidades que no pertenecen al léxico de la lengua de partida. En algunos casos estos elementos debían haberse traducido, en otros no. Bajo esta categoría también incluimos los errores en la traducción de funciones o nombres específicos de programas de software o aplicaciones, que son definidos por una terminología específica —por ejemplo la terminología de Microsoft o de la Unión Europea— y por ello tienen un carácter similar al de los nombres propios.

Nombre	Presencia de léxico en una lengua extranjera diferente a la lengua de partida.	
Número	1.a.1	
Definición	En el original se utilizan palabras en otro idioma y estas se traducen erróneamente o se traducen, cuando no se deben traducir.	
Valor del error	Mistranslation, omission, untranslated	
Ejemplo(s):	Original:	<i>"Best in Class" Gesamtwirkungsgrade basierend auf einem neuen Wilo-Trockenläuferdesign.</i>
	TA:	"Best en Class" Rendimientos completos sobre la base de un diseño de corredores seco de Wilo nuevo.
	PE:	"Best in Class" Rendimientos completos sobre la base de un nuevo diseño de rotores secos de Wilo.

Tabla 148: Definición de la categoría presencia de léxico extranjero

8. Clasificación de errores para la sistematización de la posesición

El problema de los nombres propios en nuestro corpus se ve agravado por el hecho de que algunos nombres se componen de anglicismos, cifras y/o siglas. Además no tienen ningún tipo de marcación, lo que dificulta su identificación.

Nombre	Tratamiento de nombres propios, abreviaturas o siglas	
Número	1.a.2	
Definición	En el original hay nombres propios, p.ej. nombres de productos cuyas marcas están registradas, que no son reconocidos como tales y se traducen cuando no es necesario o no se traducen correctamente. También incluimos en esta categoría los errores en la traducción de nombres de funciones específicas, como botones o funciones de un programa.	
Valor del error	Mistranslation, omission, untranslated	
Ejemplo(s):	Original:	<i>Halten Sie Apache HTTP Server an und schließen Sie ihn</i>
	TA:	mantenga su ataque HTTP servidor y cierren lo
	PE:	Detenga el servidor Apache HTTP y ciérrelo

Tabla 149: Definición de la categoría tratamiento de nombres propios

Categoría 1b: errores de naturaleza extralingüística

Bajo esta categoría agrupamos los problemas que no tienen que ver con la traducción misma de un texto sino más bien con su adaptación a una lengua o contexto, como en el caso de la conversión de las unidades de medida o moneda o la localización de enlaces e hipervínculos. En el subcorpus que utilizamos para la segunda fase del análisis no encontramos ningún error en la conversión de convenciones, sin embargo lo incluimos porque lo consideramos relevante e importante de tener en cuenta en el momento de postular directrices o ayudas para el poseedor.

Nombre	Errores en las conversión de convenciones como unidades de medida, números, moneda, etc.
Número	1.b.1
Definición	En el texto de destino no se han ajustado los valores numéricos de las diferentes unidades (por ejemplo, monedas, métricas, sistemas de medición, formatos de fecha, etc.).
Valor del error	Mistranslation, omission, addition, untranslated

Tabla 150: Definición de la subcategoría de errores en la conversión de convenciones

8. Clasificación de errores para la sistematización de la posesición

También como problemas de naturaleza extralingüística encontramos la siguiente categoría en la que incluimos los errores en la interpretación de elementos tipográficos, que no se identifican como tales y por ello no se sustituyen o no se emplean correctamente en la traducción.

Nombre	Errores de funcionalidad de elementos tipográficos	
Número	1.b.2	
Definición	Bajo esta categoría incluimos cuestiones y problemas relacionados con la interpretación errónea de elementos tipográficos como barras, asteriscos, corchetes o guiones, cuya función (por ejemplo, de conjunción, aclaración o separación de cifras o fechas) no se identifica correctamente.	
Valor del error	Addition, omission, misplaced	
Ejemplo(s):	Original:	<i>Die Optionen Netzwerkeinstellungen, Datum des Servers ändern und Server neu starten/stoppen, stehen bei der CDVersion nicht zur Verfügung</i>
	TA:	Las opciones sobreseimientos de red, fecha del servidor modifican, nuevo salen servidores, paran y no están con de reemplazo/ , de las CDVersion.
	PE:	Las opciones de configuración de red, fecha del servidor y reiniciar/parar el servidor no están disponibles en la versión de CD.

Tabla 151: Definición de la subcategoría errores de funcionalidad de elementos tipográficos

Asimismo, nos encontramos con errores en la traducción o adaptación de elementos o caracteres utilizados para la formulación de enlaces también hacen parte de los elementos extralingüísticos en los que encontramos algunos errores:

Nombre	Enlaces, vínculos, hipertexto, referencias	
Número	1.b.3	
Definición	Un texto contiene enlaces p.ej. a páginas web con características tipográficas especiales, en algunos casos no se reconocen como tales y son tratados como texto, se añaden u omiten espacios o se traducen erróneamente.	
Valor del error	Addition, omission, mistranslation, misplaced	
Ejemplo(s):	Original:	<i>Zertifikat abrufbar unter www.wilo.de/rechtliches</i>
	TA:	certificado bajo www.wilo.de / legal
	PE:	certificado disponible en www.wilo.de/información legal

Tabla 152: Definición de la subcategoría traducción errónea de enlaces, hipertexto y referencias

8. Clasificación de errores para la sistematización de la posesición

8.3.2 *Segunda macrocategoría: errores relacionados con las características propias del alemán*

En el apartado “1.5 Problemas de traducción en la combinación lingüística alemán-castellano” a partir de la página 47, describimos algunos de los puntos problemáticos que pueden ocasionar errores en la traducción. En este apartado no nos ocuparemos del alemán como sistema, sino de aquellos aspectos del alemán que pueden llevar a algún sistema de traducción automática a cometer un error. Es decir, nos referimos al alemán como sistema lingüístico, y por eso no trataremos aquí aspectos detallados de la combinación lingüística alemán-castellano, de los que nos ocuparemos en la cuarta macrocategoría (a partir de la página 301).

En el alemán encontramos ciertas características, como el uso de composiciones, que constituyen una causa frecuente de errores, particularmente en la traducción automática. No solo la traducción de sustantivos o nombres compuestos en alemán puede ser problemática, también la traducción de verbos separables, cuyos componentes se separan cuando el verbo está conjugado y que por ello dificultan el proceso de la traducción automática, ya que en algunos casos no se identifican como unidad.

Igualmente la conjugación de los verbos ocasiona dificultades en la traducción del alemán, pero este problema lo tratamos ya anteriormente (véase “Clasificación de LISA” a partir de la página 131).

Categoría 2.a: errores en la traducción de verbos separables

En alemán existe la posibilidad de añadir prefijos a los verbos dándole a estos un significado totalmente nuevo y convirtiéndolos en nuevas palabras. En algunos casos el prefijo se separa del verbo y se coloca al final de la oración. La parte principal del verbo se conjuga y se posiciona en la oración —dependiendo del tipo— en la segunda posición. Ya en la primera fase del análisis pudimos identificar que se presentan problemas en la traducción automática de estos verbos porque no se identifican como una unidad, y la parte principal y el prefijo se traducen independientemente lo que genera problemas de orden léxico.

8. Clasificación de errores para la sistematización de la posesición

Nombre	Verbos compuestos separables	
Número	2.a	
Definición	Errores que se dan cuando los verbos separables no son identificados como unidad y por ello se traducen erróneamente.	
Valor del error	Mistranslation, omission, untranslated	
Ejemplo(s):	Original:	<i>Bei einem Druck auf diese ordnen sie sich auf der linken Seite an.</i>
	TA:	En una presión a la que se inscriben en la izquierda.
	PE:	Pulsando sobre ellas se colocan en la izquierda.

Tabla 153: Definición de la subcategoría traducción de verbos compuestos separables

Categoría 2.b: errores en la traducción de conjunciones y partículas

Errores en la traducción de palabras necesarias para conectar y hacer comprensible el texto y que varían en cada idioma. La traducción del original puede reproducir una de las funciones de la conjunción o de la partícula en el texto original, pero en la traducción al castellano no se utiliza el equivalente correcto, que en algunos casos puede ser una conjunción o una partícula diferente.

Nombre	Conjunciones y partículas	
Número	2.b	
Definición	Errores en la traducción de conjunciones o partículas y de su sintagma en oraciones; aunque la traducción sea correcta no se selecciona —debido a la polisemia— el equivalente correcto para la oración en castellano.	
Valor del error	Mistranslation, omission, untranslated	
Ejemplo(s):	Original:	<i>Dazu ist auf die entsprechende Zahl zu klicken, damit die Aktualisierung nach jeweils 1 Minute, alle 5 Minuten bzw. alle 15 Minuten erfolgt.</i>
	TA:	Para ello es de la cifra correspondiente a Internet, de la actualización de los 1 minuto, todos 5 minutos o todos los 15 minutos.
	PE:	Para ello haga clic en la cifra correspondiente, para que la actualización se realice cada 1 minuto, cada 5 minutos o cada 15 minutos.

Tabla 154: Definición de la subcategoría traducción de conjunciones y partículas

8. Clasificación de errores para la sistematización de la posesición

Categoría 2.c: Errores de transferencia sintáctica de composiciones

En alemán no se hace una distinción entre compuestos nominales y composiciones, ambos son *Komposita*; nosotros hemos hecho en nuestra clasificación una distinción y denominado composiciones a las *Kompositionen* que se separan con un guion ortográfico. Esta decisión la tomamos porque los compuestos nominales son unidades que pueden aparecer en un diccionario especializado, las composiciones de palabras unidas por uno o varios guiones no. Este tipo de palabras pueden ser creadas por el autor del texto, para evitar redundancias y dar claridad a las oraciones complejas. En el corpus encontramos ejemplos de otro tipo de composiciones de palabras compuestas unidas con un guion, por ejemplo porque uno de los elementos es un extranjerismo (*E-Mail-Adresse*) o un nombre propio (*TÜV-Zertifikat*).

Nombre	Transferencia sintáctica de composiciones	
Número	2.c	
Definición	La traducción no es correcta porque el orden de los componentes no da el sentido original, porque no se traducen o se omiten algunos de los elementos de la composición.	
Valor del error	Addition, omission, mistranslation, untranslated	
Ejemplo(s):	Original:	<i>Einbau- und Betriebsanleitung</i>
	TA:	Einbau- y funcionamiento orientación
	PE:	instrucciones de montaje y funcionamiento

Tabla 155: Definición de la subcategoría error de transferencia sintáctica en composiciones

Categoría 2.d: errores en la traducción de oraciones compuestas

Pueden presentarse errores en la traducción automática de oraciones compuestas; en el corpus de nuestra tesis nos encontramos dos tipos de oraciones compuestas. Hallamos oraciones subordinadas, es decir aquellas en las que existe una relación de dependencia.

Nombre	Oraciones subordinadas
Número	2.d.1
Definición	No se traducen correctamente oraciones subordinadas sustantivas, adjetivas o adverbiales
Valor del error	Mistranslation

8. Clasificación de errores para la sistematización de la posesición

Ejemplo(s):	Original:	<i>Mit der ordnungsgemäßen Entsorgung und des sachgerechten Recycling dieses Produktes werden Umweltschäden und eine Gefährdung der persönlichen Gesundheit vermieden.</i>
	TA:	Con la eliminación adecuada y reciclado se este producto daños medioambientales y una amenaza para la salud personal.
	PE:	Con la eliminación adecuada y el reciclado correcto de este producto se evitarán daños medioambientales y los riesgos para la salud personal.

Tabla 156: Definición de la subcategoría errores en la traducción de oraciones subordinadas

También nos encontramos con errores en oraciones compuestas coordinadas, que tienen el mismo valor funcional y son sintácticamente independientes.

Nombre	Oraciones coordinadas	
Número	2.d.2	
Definición	No se reconoce el vínculo entre las oraciones y por ello no se traducen correctamente.	
Valor del error	Mistranslation	
Ejemplo(s):	Original:	<i>Installation und elektrischen Anschluss nur durch Fachpersonal und gemäß geltenden Vorschriften durchführen lassen</i>
	TA:	Instalación y conexión eléctrica solo por personal especializado y llevar a cabo de conformidad con las normas en vigor!
	PE:	La instalación y la conexión eléctrica solo debe llevarlas a cabo personal especializado y de conformidad con las normas en vigor!

Tabla 157: Definición de la subcategoría errores en la traducción de oraciones coordinadas

Categoría 2.e: errores en la traducción de negaciones

Como vimos en la primera parte (véase página 68) la traducción de la negación puede ser problemática, ya que si la estructura sintáctica es compleja se puede interpretar erróneamente lo que se desea negar.

Nombre	Negación
Número	2.e
Definición	No se transfiere correctamente el sentido de la negación.
Valor del error	Mistranslation, addition, omission, untranslated

8. Clasificación de errores para la sistematización de la posesición

Ejemplo(s):	Original:	<i>Das Gerät darf keinen Temperaturen außerhalb der Bereiche -10 °C bis +50 °C ausgesetzt werden.</i>
	TA:	El aparato no puede ser expuesto ninguna temperaturas fuera de los ámbitos -10 hasta +50 °C.
	PE:	El aparato no puede ser expuesto a temperaturas fuera de los ámbitos -10 °C hasta +50 °C .

Tabla 158: Definición de la subcategoría errores en la traducción de negaciones

Categoría 2.f: errores en la traducción de colocaciones o fraseologías

Errores en la traducción de expresiones o fraseologías en las que sus elementos tienen un significado especial e incluso diferente del significado si se analizan por separado (véase página 60). Tanto las fraseologías como las colocaciones son de uso común y no de terminología especializada.

Nombre	Colocaciones o fraseologías	
Número	2.f	
Definición	No se traducen correctamente colocaciones o fraseologías.	
Valor del error	Mistranslation, untranslated, misplaced	
Ejemplo(s):	Original:	<i>Schwarz auf weiß.</i>
	TA:	Por escrito.
	PE:	Hablando claro.

Tabla 159: Definición de la subcategoría errores en la traducción de colocaciones o fraseologías

8.3.3 Tercera macrocategoría: errores relacionados con las características del castellano

Bajo esta macrocategoría clasificamos los errores que se presentan cuando a pesar de haberse hecho una traducción correcta del alemán al castellano, las traducciones no se ajustan a las reglas gramaticales del idioma meta. Aquí encontramos los errores que se dan cuando no hay, por ejemplo, concordancia en la flexión de género o número o en la flexión verbal. También problemas de estilo u ortotipográficos. En esta última categoría incluimos los errores de puntuación, uso incorrecto de mayúsculas y minúsculas y signos diacríticos.

Otros problemas que incluimos en esta categoría son los errores que se dieron en la sintaxis, por ejemplo, cuando se reproduce la estructura sintáctica del alemán y no se adapta a

8. Clasificación de errores para la sistematización de la posesición

la estructura gramatical del castellano. En el primer análisis marcamos algunos de estos casos como errores en el orden gramatical.

Categoría 3.a: errores de flexión o concordancia nominal

Errores de flexión o concordancia nominal se dan cuando dos o más palabras no concuerdan con respecto las características gramaticales género y número. Este tipo de error se da en los casos en que, por ejemplo, las palabras en alemán, particularmente los sustantivos, se escriben igual en singular y plural o no tienen diferencia de género. Sin embargo, en castellano si existen estas diferencias y por ello se presentan errores cuando el motor de traducción no escoge el equivalente correcto.

Nombre	Género	
Número	3.a.1	
Definición	Falta de concordancia de género entre dos o más elementos de la frase.	
Valor del error	Mistranslation	
Ejemplo(s):	Original:	<i>AKO-5004 Bedienungsanleitung der software</i>
	TA:	AKO-5004 manual de instrucciones de la software
	PE:	AKO-5004 Manual de instrucciones del software

Tabla 160: Definición de la subcategoría errores en la flexión de género

También se dan casos, como en el ejemplo a continuación, en que a pesar de que existe una diferencia entre el plural y el singular en alemán, el sustantivo alemán en singular es traducido erróneamente por el equivalente plural en castellano.

Nombre	Número	
Número	3.a.2	
Definición	Falta de concordancia de número entre dos o más elementos de la frase.	
Valor del error	Mistranslation	
Ejemplo(s):	Original:	<i>klicken Sie auf die Schaltfläche</i>
	TA:	hagan un clic en los botones
	PE:	haga clic en el botón

Tabla 161: Definición de la subcategoría errores en la flexión de número

8. Clasificación de errores para la sistematización de la posesición

Categoría 3.b: errores de flexión verbal

Ya en la primera fase del análisis pudimos ver que se en nuestro corpus se presentan errores en la flexión de los verbos. En este tipo de errores pudimos identificar varias subcategorías. En primer lugar encontramos errores en la conjugación del tiempo verbal:

Nombre	Tiempo	
Número	3.b.1	
Definición	En una forma verbal no se conjuga el tiempo verbal correctamente.	
Valor del error	Mistranslation	
Ejemplo(s):	Original:	<i>Aktualisierte Informationen finden Sie auf unserer Webseite</i>
	TA:	Encuentra informaciones actualizadas en nuestra página web
	PE:	Encontrará informaciones actualizadas en nuestra página web

Tabla 162: Definición de la subcategoría errores en la flexión verbal (tiempo)

También encontramos segmentos en los que no se identificó ni tradujo correctamente el modo, por ejemplo, porque en alemán la forma del verbo infinitivo, es igual al imperativo de este o a la tercera persona plural del modo indicativo.

Nombre	Modo	
Número	3.b.2	
Definición	En una forma verbal no se conjuga el modo correctamente.	
Valor del error	Mistranslation	
Ejemplo(s):	Original:	<i>Automatisch aktualisieren.</i>
	TA:	Automáticamente actualizan.
	PE:	Actualizar automáticamente.

Tabla 163: Definición de la subcategoría errores en la flexión verbal (modo)

Además encontramos errores en la flexión verbal, cuando no se conjuga la persona correctamente, por ejemplo, cuando se confunde la forma de tratamiento formal en alemán con la tercera persona plural.

8. Clasificación de errores para la sistematización de la posesición

Nombre	Persona	
Número	3.b.3	
Definición	En una forma verbal no se conjuga la persona correctamente.	
Valor del error	Mistranslation	
Ejemplo(s):	Original:	<i>Anschließend klicken Sie auf Anwenden:</i>
	TA:	A continuación, hagan un clic en aplicar:
	PE:	A continuación, haga clic en "aplicar":

Tabla 164: Definición de la subcategoría errores en la flexión verbal (persona)

En alemán se utiliza más frecuentemente la voz pasiva en textos técnicos que en castellano. Independientemente de esto nos encontramos casos en que la traducción automática utilizó la voz pasiva, mientras que en el original se emplea la voz activa:

Nombre	Voz	
Número	3.b.4	
Definición	No se utiliza la voz gramatical correcta para el contexto (activa o pasiva)	
Valor del error	Mistranslation	
Ejemplo(s):	Original:	<i>Durch Klicken auf das Symbol erscheinen vier Registerkarten:</i>
	TA:	Por hacer clic en el símbolo se publican cuatro tarjetas de registro:
	PE:	Al hacer clic en el símbolo aparecen cuatro fichas de registro:

Tabla 165: Definición de la subcategoría errores en la flexión verbal (voz)

Categoría 3.c: errores de registro y estilo

Nombre	Estilo	
Número	3.c	
Definición	El texto tiene problemas estilísticos: está escrito con un estilo torpe; utiliza un nivel de formalidad alta o más baja que la requerida por las especificaciones del lenguaje en general (registro) o el contenido es gramaticalmente correcto pero no idiomático.	
Valor del error	Register, not idiomatic	
Ejemplo(s):	Original:	<i>Im Einzelnen kann Nichtbeachtung beispielsweise folgende Gefährdungen nach sich ziehen:</i>

8. Clasificación de errores para la sistematización de la posesición

	TA:	Según cada caso la inobservancia puede arrastrar peligros por ejemplo siguientes:
	PE:	Según cada caso la inobservancia puede arrastrar peligros como por ejemplo los siguientes:

Tabla 166: Definición de la subcategoría errores de estilo

Categoría 3.d: errores ortotipográficos

En esta categoría agrupamos todos los errores, incluyendo los errores de puntuación, que tienen que ver con el uso y convenciones que rigen la escritura por medios tipográficos.

Nombre	Errores ortográficos	
Número	3.d.1	
Definición	Errores de cambio de letras, unión y/o separación de palabras, uso de mayúsculas y minúsculas y signos diacríticos.	
Valor del error	Omission, addition, misplaced, questionable	
Ejemplo(s):	Original:	<i>Neue Chancen mit neuen Regeln / ErP-Direktive - 11/2011</i>
	TA:	Nuevas oportunidades con nuevas normas / ErP-directiva - 11 / 2011
	PE:	Nuevas oportunidades con nuevas normas / Directiva EuP - 11/2011

Tabla 167: Definición de la subcategoría errores ortográficos

Separadamente marcamos los errores tipográficos:

Nombre	Errores tipográficos	
Número	3.d.2	
Definición	Las cuestiones relacionadas con la presentación del texto, como la utilización de espacios en blanco, guiones, paréntesis, comillas o corchetes.	
Valor del error	Omission, addition, mistranslation, questionable	
Ejemplo(s):	Original:	<i>Wenn Sie auf Registriergerät klicken, öffnet sich folgendes Fenster:</i>
	TA:	Si hace sobre registrador clic, se abre ventana siguiente:
	PE:	Si hace sobre clic "equipo de registro", se abre la ventana siguiente:

Tabla 168: Definición de la subcategoría errores tipográficos

8. Clasificación de errores para la sistematización de la posesición

Categoría 3.e: errores de sintaxis

Aquí agrupamos los errores que se dieron cuando se infringen las reglas para la construcción de sintagmas y oraciones en castellano.

Nombre	Errores de sintaxis	
Número	3.e	
Definición	Errores relacionados con la sintaxis del texto, que no tienen que ver con la ortografía, la flexión nominal o verbal, que ya se definen en otras categorías.	
Valor del error	Addition, misplaced, omission	
Ejemplo(s):	Original:	<i>Nutzen Sie also jetzt Ihren Wissens- und Kompetenzvorsprung.</i>
	TA:	Utilice su ventaja de conocimientos y de competencia pues ahora .
	PE:	Utilice pues ahora las ventajas de su conocimiento y competencia.

Tabla 169: Definición de la subcategoría errores de sintaxis

8.3.4 Cuarta macrocategoría: errores en la microestructura

Encontramos también problemas a nivel de la microestructura de los segmentos, es decir errores que se producen por impresiones en la traducción de palabras aisladas o expresiones de más de una palabra, con o sin carga semántica. Estos son el tipo de errores que marcamos en el primer análisis como uso incorrecto de preposiciones o traducción errónea de artículos y/o adverbios; también los errores en la traducción de palabras con carga semántica como sustantivos, adjetivos y verbos.

Categoría 4.a: errores en la traducción de palabras sin carga semántica

Aquí nos encontramos con las palabras que no son portadoras de contenido, cuya función es determinar o relacionar las palabras en una oración que si tienen contenido o carga semántica. Entre estas palabras encontramos que son frecuentes los errores en la traducción de preposiciones, ya que aunque se traduzcan correctamente en castellano no necesariamente se emplea la misma preposición; además en otros casos el motor de traducción emplea un equivalente inapropiado en el segmento.

Nombre	Preposiciones
Número	4.a.1.1

8. Clasificación de errores para la sistematización de la posesición

Definición	Uso inapropiado de preposiciones aisladas.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>maximale Förderhöhe in m bei $Q = 0 \text{ m}^3/\text{h}$</i>
	TA:	altura de fomentar máxima en m en $Q = 0 \text{ m}^3/\text{h}$
	PE:	altura de impulsión máxima en m para $Q = 0 \text{ m}^3/\text{h}$

Tabla 170: Definición de la subcategoría errores en la traducción de preposiciones

No solo las preposiciones aisladas generan problemas; también nos encontramos con errores en la traducción de sintagmas o grupos preposicionales.

Nombre	Sintagmas preposicionales	
Número	4.a.1.2	
Definición	Traducción incorrecta de sintagmas preposicionales.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>Möglicherweise wird im Rahmen des Setupvorgangs die folgende Nachricht angezeigt</i>
	TA:	Aquellos posiblemente al amparo de es señalado del suceso de setup noticia siguiente.
	PE:	Es posible que durante el proceso de instalación se muestre el siguiente mensaje:

Tabla 171: Definición de la subcategoría errores en la traducción de sintagmas preposicionales

En la siguiente categoría agrupamos los errores que tienen que ver con la traducción de artículos, pronombres, etc. Aquí pudimos ver que la traducción no se ajustó en algunas ocasiones al género o número de la palabra determinada; además encontramos casos en que se añadieron u omitieron determinantes.

Nombre	Determinantes	
Número	4.a.2	
Definición	Errores en la traducción de palabras que determinan un sintagma nominal como artículos, pronombres demostrativos, posesivos o interrogativos, cuantificadores, predeterminantes	
Valor del error	Mistranslation, addition, omission, untranslated	

8. Clasificación de errores para la sistematización de la posesición

Ejemplo(s):	Original:	<i>Bedienung und Konfiguration über bekannte "Roter Knopf"-Technologie</i>
	TA:	Servicio y configuración sobre tecnología de botón rojos
	PE:	Mando y configuración mediante la tecnología de “botón rojo”

Tabla 172: Definición de la subcategoría errores en la traducción de determinantes

En la traducción de adverbios, palabras sin carga semántica que complementan un verbo, un adjetivo y en algunos casos hasta oraciones, también se presentan errores porque no se escoge el equivalente correcto y como en el caso de las preposiciones, en algunos casos se añaden o se omiten sus traducciones.

Nombre	Adverbios	
Número	4.a.3.1	
Definición	Traducción errónea de adverbios que modifiquen tanto verbos, como adjetivos u otros adverbios.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>Danach klicken Sie auf Annehmen</i>
	TA:	A eso hace sobre aceptar clic
	PE:	Después haga clic sobre “Aceptar”

Tabla 173: Definición de la subcategoría errores en la traducción de adverbios

Del mismo modo que en el caso de las preposiciones nos encontramos con errores en las traducciones de algunas estructuras adverbiales, es decir aquellos casos en que el adverbio no puede traducirse aisladamente:

Nombre	Estructuras adverbiales	
Número	4.a.3.2	
Definición	Traducción errónea de estructuras adverbiales.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>Zurück zum Wizard</i>
	TA:	De vuelta a la Wizard
	PE:	Volver al asistente

Tabla 174: Definición de la subcategoría errores en la traducción de estructuras adverbiales

8. Clasificación de errores para la sistematización de la posesición

También nos encontramos con errores en las traducciones de ciertas conjunciones y partículas. En esta categoría incluimos los errores que se dan cuando se añade o se omite una conjunción o partícula en castellano y por ello —aunque tenga el mismo nombre— los errores aquí marcados no son los mismos que los que describimos en la categoría 2.b, en la que nos referimos a los errores de traducción de conjunciones y partículas alemanas (véase p. 293).

Nombre	Conjunciones y partículas	
Número	4.a.4	
Definición	Traducción errónea de conjunciones y partículas.	
Valor del error	Adition, omission	
Ejemplo(s):	Original:	<i>Schließlich meldet die Anwendung, dass die Deinstallation abgeschlossen worden ist</i>
	TA:	Por último, la aplicación que revela que la Deinstallation ha concluido.
	PE:	Por último, la aplicación indica que la desinstalación ha concluido.

Tabla 175: Definición de la subcategoría errores en la traducción de conjunciones y partículas

Categoría 4.b: errores en la traducción de palabras con carga semántica

Bajo esta categoría anotamos todos los errores que se presentaron en la traducción de palabras que tienen carga semántica autónoma, es decir sustantivos, adjetivos y verbos. En estos casos la traducción fue errónea porque no se encontró el equivalente adecuado o porque no se tradujo la palabra. También nos encontramos algunos casos en que verbos fueron traducidos como sustantivos porque estaba en mayúsculas.

Nombre	Sustantivos	
Número	4.b.1	
Definición	Errores en la traducción de sustantivos.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>Daraufhin werden zwei weitere Nachrichten angezeigt</i>
	TA:	Después se indican dos otras noticias
	PE:	Después se indican otros dos mensajes

Tabla 176: Definición de la subcategoría errores en la traducción de sustantivos

8. Clasificación de errores para la sistematización de la posesición

En la siguiente subcategoría clasificamos la traducción errónea de adjetivos que no tiene que ver con la concordancia de género o número.

Nombre	Adjetivos	
Número	4.b.2	
Definición	Errores en la traducción de adjetivos.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>Einsparpotenzial bei Trockenläufern weltweit</i>
	TA:	Potencial de ahorrar en corredores secos universal
	PE:	Potencial de ahorro en rotores secos a nivel mundial

Tabla 177: Definición de la subcategoría errores en la traducción de adjetivos

También incluimos errores en la traducción de verbos, que se dieron porque no se encontró el equivalente correcto en castellano, porque se añadieron o se omitieron, o porque se dejaron sin traducir. Aquí hicimos una distinción entre los errores presentados en la traducción de verbos independientes y en la traducción de formas verbales compuestas.

Nombre	Verbos simples	
Número	4.b.3.1	
Definición	Errores en la traducción de verbos.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>Deinstallieren Sie Apache HTTP Server 2.2.3</i>
	TA:	Deinstallieren usted ataque HTTP servidor 2.2.3.
	PE:	Desinstale el Servidor HTTP Apache 2.2.3.

Tabla 178: Definición de la subcategoría errores en la traducción de verbos

Aunque en muchas ocasiones se traducen los verbos aislados correctamente, pueden presentarse dificultades si en la oración se encuentran formas verbales compuestas como el perfecto alemán.

Nombre	Formas verbales compuestas
Número	4.b.3.2

8. Clasificación de errores para la sistematización de la posesición

Definición	Errores en la traducción de verbos.	
Valor del error	Mistranslation, addition, omission, untranslated, misplaced	
Ejemplo(s):	Original:	<i>Ab jetzt installiert das System nacheinander verschiedene Produkte, die an AKO 5004 beteiligt sind.</i>
	TA:	A partir de ahora el sistema instala los unos a los otros productos diferentes que están hechos en AKO 5004 participar.
	PE:	A partir de ahora el sistema instala uno tras otro diferentes productos que inter-vienen en AKO 5004.

Tabla 179: Error en la traducción de formas verbales compuestas

Como indicamos al hablar de la segunda macrocategoría en la que agrupamos los errores relacionados con las características propias del alemán, también nos encontramos con errores de transferencia sintáctica y de fluidez en compuestos nominales. Aunque los problemas que pueden generar la traducción de los compuestos nominales en alemán pueden considerarse como problemas típicos del alemán, decidimos analizarlos como errores propios de la microestructura en la que analizamos los aspectos de la combinación lingüística alemán castellano. Ello debido a que la traducción de los compuestos nominales no presenta los mismos problemas en todas las combinaciones lingüísticas, en idiomas con una estructura sintáctica más similar a la del alemán.

La utilización de compuestos nominales es más usual en los textos técnicos en alemán que en castellano. Para este tipo de palabras no hay una palabra equivalente en castellano, generalmente en castellano dan lugar a un sintagma. En algunos casos el compuesto nominal se traduce erróneamente y aunque se escogen los elementos aislados correctamente, sintácticamente no se reproduce la relación implícita en el original.

Nombre	Error de transferencia sintáctica de compuestos nominales
Número	4.b.4.1
Definición	Se traducen correctamente los componentes pero el orden no es correcto y por ello la traducción tampoco lo es. También se incluyen aquí los errores que se dan cuando las traducciones de los elementos del compuesto no se relacionan coherentemente (también por omisión de componentes) o cuando faltan elementos.
Valor del error	Mistranslation, addition, omission, untranslated

8. Clasificación de errores para la sistematización de la posesición

Ejemplo(s):	Original:	<i>Bei einem Doppelklick auf den Gerätenamen erhalten Sie Zugang zu ihrem Parametersatz:</i>
	TA:	En un doble clic o a los aparatos nombre tendrá acceso a su parámetros frase:
	PE:	Haciendo doble clic en el nombre de los aparatos tendrá acceso a su conjunto de parámetros:

Tabla 180: Definición de la subcategoría error de transferencia sintáctica en compuestos nominales

También nos encontramos con casos en que los componentes se traducen aisladamente, y no como una palabra, lo que lleva a que el orden de los componentes en la traducción no sea sintácticamente apropiado.

Nombre	Error en la fluidez de los sintagma	
Número	4.b.4.2	
Definición	No existe siempre una palabra equivalente en castellano para cada compuesto nominal o composición en alemán. Por ello la traducción es generalmente un sintagma. En algunos casos se traducen correctamente los componentes pero la fluidez, por ejemplo por problemas de género o número, no es correcta.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>Kommunikationsfähig für die Gebäudeautomation durch nachrüstbare IF-Module.</i>
	TA:	Apto para Comunicación para la automatización de edificio por módulos de IF nachrüstbare.
	PE:	Apto para la comunicación para la automatización de edificios por módulos de IF actualizables.

Tabla 181: Definición de la subcategoría error de transferencia sintáctica léxica en los compuestos nominales

Adicionalmente se presentan errores en la traducción del significado de un compuesto o una composición; en algunos casos se traducen aisladamente sus componentes o se escoge erróneamente el equivalente. En esta categoría incluimos los errores en la traducción de sustantivos compuestos integrados por al menos dos morfemas o construcciones de morfemas independientes. También los errores en la traducción de compuestos de sustantivos unidos por un guion y que denominaremos composiciones, ya que alguno de los componentes puede ser una letra, una abreviatura, una cifra o un nombre propio. En algunos casos se utiliza el guion

8. Clasificación de errores para la sistematización de la posesición

para destacar alguno de los constituyentes o hacer más claras las composiciones múltiples (Oster 2003, 44).

Nombre	Transferencia léxica de compuestos nominales y composiciones	
Número	4.b.5	
Definición	Errores en la traducción de compuestos nominales y composiciones.	
Valor del error	Mistranslation, addition, omission, untranslated	
Ejemplo(s):	Original:	<i>Einsparpotenzial bei Nassläufern in Europa:</i>
	TA:	Potencial de ahorrar en corredores mojados en Europa:
	PE:	Potencial de ahorro en rotores húmedos en Europa:

Tabla 182: Error en la transferencia léxica de compuestos nominales y composiciones

8.3.5 Quinta macrocategoría: errores relacionados con el género y funcionalidad del texto o la redacción del original

Como ya lo indicamos en el marco teórico (p.187) los problemas o errores en los textos originales tienen un papel importante en la traducción automática ya que pueden ocasionar errores en la traducción. Nos referimos al tipo de problemas que surgen por defectos en el original, que pueden ser de tipo formal (p.ej. repeticiones erróneas de una palabra, omisiones o referencias incorrectas) que no pueden ser detectados por los motores de traducción automática. También nos encontramos aquí los problemas de traducción que se pueden dar por errores de ortografía o errores de imprenta.

En esta macrocategoría incluimos también los errores de redacción del texto y los problemas concretos que se encontraron en los segmentos originales de nuestro corpus. Nos referimos primordialmente aquí a la presencia de etiquetas que pueden ser interpretadas erróneamente como texto.

Categoría 5.a: errores por manejo erróneo de las etiquetas

Nombre	Etiquetas en el original
Número	5.a
Definición	Errores relacionados con etiquetas (códigos usados para representar la estructura o el formato del texto, también conocidos como “tags”).

8. Clasificación de errores para la sistematización de la posesición

Valor del error	Omission, misplaced, questionable	
Ejemplo(s):	Original:	<ut/>Für den ordnungsgemäßen Betrieb
	TA:	Las cosas ut / me para el buen funcionamiento de AKO5004
	PE:	<ut/>Para el buen funcionamiento de AKO5004

Tabla 183: Definición de la subcategoría errores causados por la presencia de etiquetas en el original

Categoría 5.b: errores de redacción en el texto original

En los casos que marcamos estos errores, la redacción del original era confusa o errónea. En el ejemplo a continuación tenemos cuatro sustantivos: *Energiekosten*, *Pumpenempfehlung*, *Energiekosten* y *Heizungspumpe* y un adjetivo *ungeregelte* que califica al sustantivo *Heizungspump*, pero las palabras no tienen relación, ni conexión alguna.

Nombre	Registro/estilo	
Número	5.b	
Definición	Problemas de registro o estilo en el original	
Valor del error	not idiomatic, register	
Ejemplo(s):	Original:	<i>Energiekosten Pumpenempfehlung Energiekosten ungeregelte Heizungspumpe</i>
	TA:	Costes de energía recomendación de bomba costes de energía Heizungspumpe desorganizado
	PE:	Costes de energía recomendación de bomba costes de energía bomba de calefacción no regulada

Tabla 184: Definición de la subcategoría errores causados por la redacción del original

Categoría 5.c: errores ortotipográficos en el texto original

Como su nombre lo indica con esta categoría clasificamos todos los errores ortotipográficos detectados en el original.

Nombre	Error ortotipográfico	
Número	5.c	
Definición	Errores orto tipográficos y de puntuación en el original	
Valor del error	Addition, mistranslation, omission, questionable	
Ejemplo(s):	Original:	<i>Optionale Schnittstellen zur Buskommuni-kation durch einsteckbare IF-Module</i>

8. Clasificación de errores para la sistematización de la posesición

	TA:	Interfaces opcionales a la Autobús venir uni-kation por un módulos de IF enchufables
	PE:	Interfaces opcionales para comunicación de bus por medio de módulos de IF enchufables

Tabla 185: Definición de la subcategoría errores causados por errores orto-tipográficos en el original

8.3.6 Sexta macrocategoría: otro tipo de errores

Finalmente dejamos una categoría abierta para poder clasificar los errores que no se ajustaran a ninguna de las categorías que mencionamos anteriormente. Como en el ejemplo a continuación, en algunos casos se dan errores de diversos tipos, que pueden clasificarse en diferentes categorías.

Nombre	Error de otro tipo	
Número	6	
Definición	Cualquier tipo de error que no se ajusta a las categorías mencionadas anteriormente.	
Valor del error	Addition, mistranslation, omission, questionable	
Ejemplo(s):	Original:	Erst 2013, dann kann ich ja weitersehen.
	TA:	Sólo 2013, entonces puedo ver.
	PE:	Primero 2013, entonces podré seguir decidiendo.

Tabla 186: Otro tipo de errores

Una vez realizado el análisis cualitativo y la clasificación de los errores encontrados utilizando las categorías y subcategorías de errores explicadas anteriormente, hicimos el conteo de errores. En la primera columna aparece la categoría y/o subcategoría concatenada con el tipo de error y en las columnas B a E el número de errores que los poseedores editaron en cada segmento según el motor correspondiente. En la columna F aparece el número total de errores encontrados por categoría, sin tener en cuenta el motor utilizado. La captura de pantalla siguiente muestra una parte de esta tabla. La hoja de cálculo se encuentra en el CD-ROM, allí hicimos diferentes cálculos en hojas diferentes, para poder calcular y cruzar los datos analizados.

8. Clasificación de errores para la sistematización de la posesición

	A	B	C	D	E	F
1	Categoría de error	smt1	smt2	rbmt1	rbmt2	Total
2	1.a.1: mistranslation	5	5	11	1	22
3	1.a.1: untranslated		2	5	3	10
4	1.a.2: mistranslation	8	1	11	3	23
5	1.a.2: untranslated	2	5	7		14
6	1.b.1: mistranslation		2	3		5
7	1.b.2: addition	2	9	2	1	14
8	1.b.2: misplaced	1	1	1	3	6

Ilustración 47: Tabla para el conteo de errores den la segunda fase

Con esta clasificación estamos haciendo una propuesta basada en la descripción de problemas, teniendo en cuenta cómo y cuándo se dan, con cuyo análisis pretendemos formular estrategias para la formación de traductores. Habiendo explicado cada categoría entraremos ahora a describir los errores que encontramos en esta segunda fase del análisis.

8.4 Resultados globales según el tipo de motor

En esta fase pudimos marcar cuántos segmentos de los seleccionados para la posesición no fueron cambiados en la posesición; de los 651 segmentos analizados, 150 no fueron poseditados, y por ello los marcamos como correctos. De estos segmentos sin errores 24 fueron propuestas del SMT1, 14 del SMT2, 103 del RBMT1 y 8 del RBMT2. Estos datos se recogen de manera relativa en la siguiente ilustración.

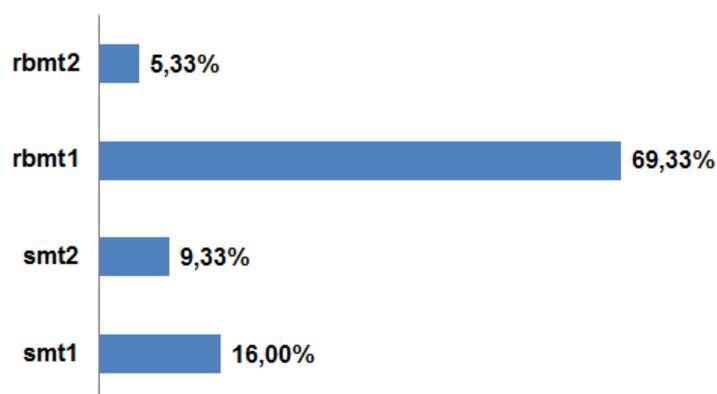


Ilustración 48: Porcentaje de propuestas marcadas como correctas según el motor

De todos los 651 segmentos analizados 111 eran propuestas realizadas por el motor SMT1, 78 del motor SMT2, 411 del RBMT1 y 51 propuestas del motor RBMT2. Ambos resultados, es decir el número de propuestas calificadas por los poseedores como correctas y el volumen de propuestas seleccionadas según el motor, confirman —como lo vimos ya en la

8. Clasificación de errores para la sistematización de la posesición

primera fase— que las propuestas del motor RBMT1 fueron favorecidas por los poseedores. Estos datos se recogen de manera relativa en la tabla siguiente:

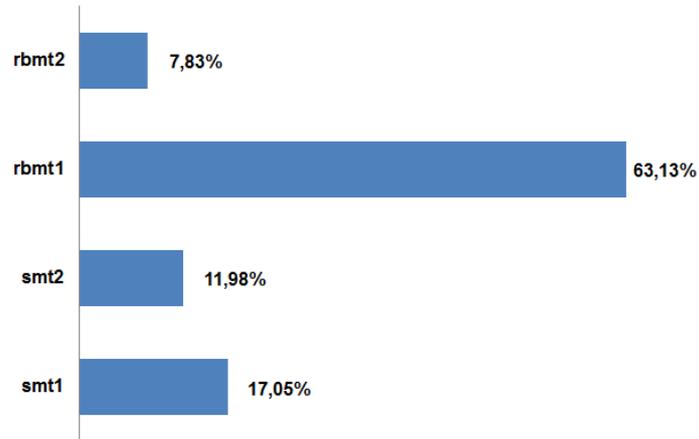


Ilustración 49: Porcentaje de propuestas seleccionadas para la posesición según el motor

En relación con el número de errores presentes en cada propuesta de TA, también pudimos contar el número de errores que se marcaron por segmento y con qué sistema se hizo la traducción respectivamente.

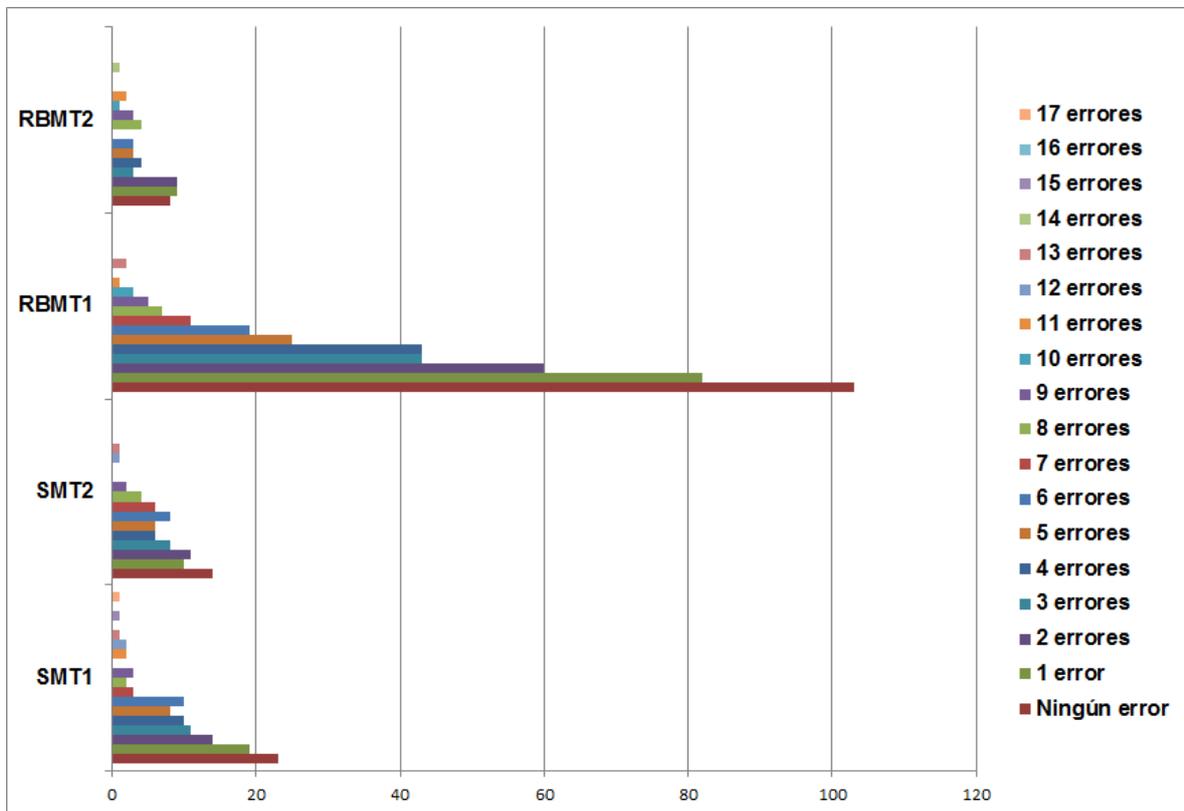


Ilustración 50: Número de errores encontrados en relación con los motores seleccionados

8. Clasificación de errores para la sistematización de la posesición

Como podemos ver en la tabla anterior el RBMT1 también es el motor en el que encontramos más segmentos con pocos errores; aunque este cálculo es relativo, porque también es el motor con más propuestas seleccionadas para la posesición. Pudimos también establecer cuáles fueron los errores más frecuentes en los segmentos en los que solo marcamos un error, sin tener en cuenta el motor como se puede ver en la tabla a continuación:

El porcentaje en que se presentaron estos errores con respecto al total de errores marcados se puede ver en la ilustración a continuación:

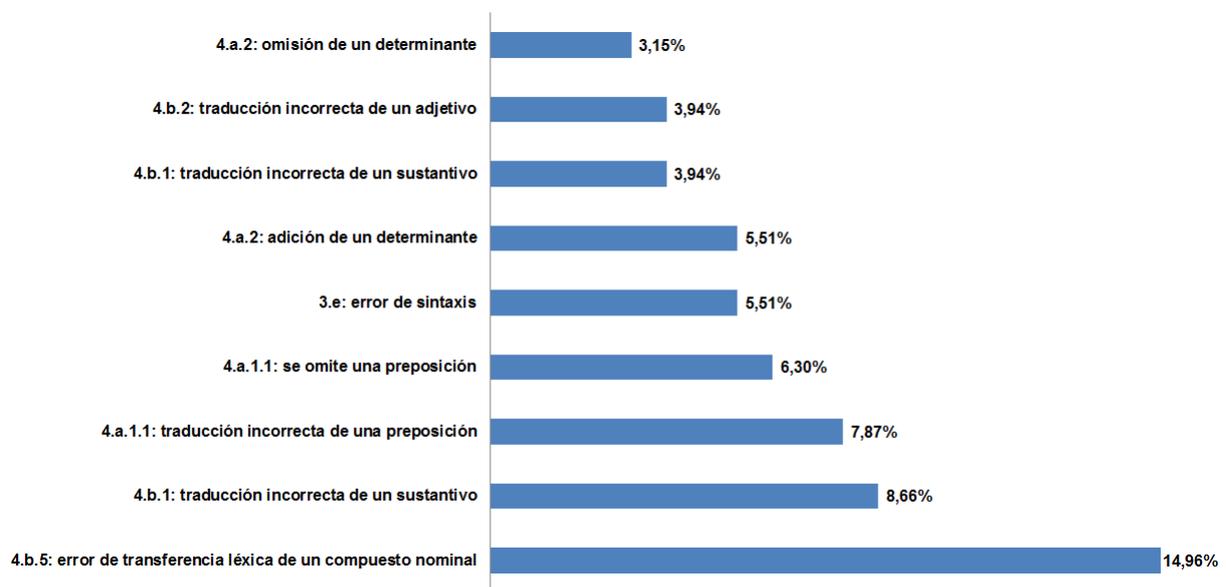


Ilustración 51: Porcentaje de errores únicos más frecuentes sin tener en cuenta el motor

De los 83 errores aislados marcados en las propuestas del RBMT1 18 son errores de transferencia léxica de un compuesto nominal, 9 traducción errónea de un sustantivo, 7 traducción errónea de una preposición, 5 traducción incorrecta de un adjetivo y 5 también omisión de una preposición.

8. Clasificación de errores para la sistematización de la posesición

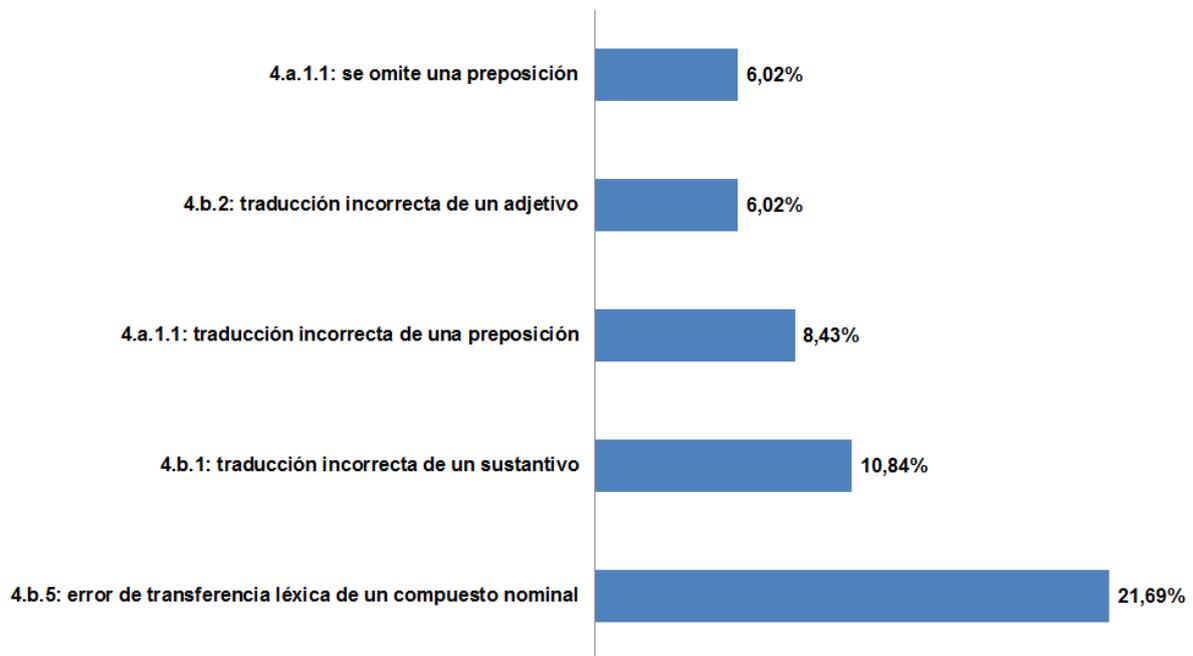


Ilustración 52: Porcentaje de errores únicos más frecuentes en las propuestas del motor rbmt1

De los 19 segmentos con solo un error en las propuestas del SMT1 en 4 ocasiones se añadieron determinantes y en 2 ocasiones se tradujo erróneamente una preposición. Los tipos de errores marcados en los segmentos restantes solo se presentaron una vez. Allí encontramos, por ejemplo, la traducción errónea de una forma verbal compuesta, tratamiento erróneo de las etiquetas, la traducción incorrecta de un sustantivo o un error de sintaxis.

En las 10 propuestas del motor SMT2 que solo tuvieron un error encontramos 2 casos en los que se añade un determinante, un error de tratamiento erróneo de etiquetas, no se traduce correctamente una preposición y en otra ocasión se omite una preposición. También encontramos un caso en el que no se interpreta correctamente un elemento tipográfico y por ello se coloca en la posición incorrecta, un error de puntuación, no se traduce una sigla y falta la traducción de una forma verbal compuesta.

Los motores estadísticos se añaden innecesariamente determinantes y hay problemas con la traducción de preposiciones. A diferencia de los motores basados en reglas, solo en las propuestas de los motores estadísticos se dañan las etiquetas porque no se interpretan correctamente.

8. Clasificación de errores para la sistematización de la posesición

En 9 segmentos de las propuestas del rbmt2 solo marcamos un error. Aquí encontramos 2 errores de sintaxis, un error en la conjugación del modo verbal, la omisión de una preposición, un error ortográfico por uso inadecuado de mayúsculas/minúsculas, no se traduce un anglicismo, además no se traduce correctamente un sustantivo y se omite un artículo.

Solo en un caso se escogió para la posesición una propuesta que no se tradujo. En el archivo de cálculo original pudimos ver que había para este segmento tanto una propuesta de la memoria de traducción y como propuestas de los otros motores de traducción. Sin embargo el poseedor escogió la propuesta del motor rbmt1 a pesar de que con este no tradujo el segmento original, porque —al parecer— le resultó más fácil traducir de cero el segmento que corregirlo.

Source	<ut/>Legende anzeigen/ausblenden<ut/> <ut/>:
tm	Ver propiedades del dispositivo.
smt1	Las cosas ut / me leyenda / encandilar suceda ut / me suceda ut / me:
smt2	Sucedá ut /. / denuncian leyenda encandilar impliquemos suceda. / ut ut /:.
rbmt1	<ut/>Legende anzeigen/ausblenden<ut/> <ut/>:
rbmt2	<ut/>Poner señalan/esconder<ut/><ut/>:
PE	<ut/>Mostrar/ocultar leyenda<ut/> <ut/>:

Tabla 187: Segmento que se tradujo de cero

También pudimos hacer un análisis estadístico de los errores que más se marcaron, independientemente del motor seleccionado. Allí nos encontramos con que el mayor problema fue el orden sintáctico. Es decir, un error que tiene que ver con las características propias del castellano, que no fueron aplicadas correctamente por el motor de traducción. Allí se marcaron en total 189 errores. A continuación aparecen los errores en la traducción de sustantivos, verbos y preposiciones. También fueron frecuentes los errores de transferencia léxica de los compuestos nominales alemanes y errores en la conjugación del modo verbal.

8. Clasificación de errores para la sistematización de la posesición

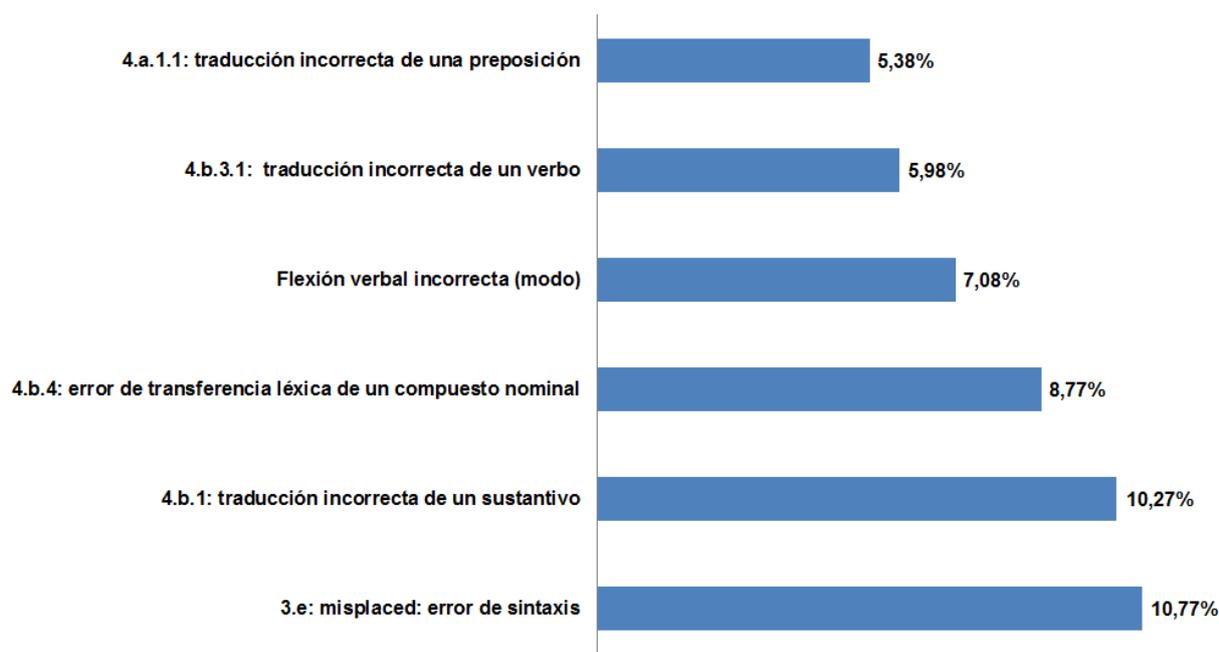


Ilustración 53: Errores más frecuentes independientemente del motor seleccionado

En la tabla a continuación daremos ejemplos de los errores e indicaremos el número de errores marcados. En algunos de los ejemplos se presentan varios errores, por ello marcaremos en color el caso a que nos referimos concretamente.

Categoría de error	Número	Ejemplo
Error de sintaxis (3.e: misplaced)	176	[TA]: <ut/>Está estandarizado el español como idioma seleccionado. [Original]: <ut/>Standardmäßig ist Spanisch als Sprache ausgewählt. [PE]: <ut/>El español está seleccionado como idioma de manera pre-determinada.
Traducción incorrecta de un sustantivo (4.b.1: mistranslation)	151	[TA]: <ut/>Debajo de esta tarjeta de registro encuentra además los parámetros en lo que se refiere a la configuración de red: [Original]: <ut/>Unter dieser Registerkarte finden Sie außerdem die Parameter bezüglich der Netzkonfiguration: [PE]: <ut/>En esta ficha de registro encontrará además los parámetros en lo que se refiere a la configuración de red:
Error de transferencia léxica de un compuesto nominal (4.b.4: mistranslation)	116	[TA]: Punto de empresa Q 13,2 metros cúbicos, H 7,6 metros [Original]: Betriebspunkt Q 13,2 m ³ , H 7,6 m [PE]: Punto de funcionamiento Q 13,2 metros cúbicos, H 7,6 metros

8. Clasificación de errores para la sistematización de la posesición

Categoría de error	Número	Ejemplo
Flexión verbal incorrecta (modo) (3.b.2: mistranslation)	105	[TA]: Desinstalar PostgreSQL 8.1. [Original]: <i>Deinstallieren Sie PostgreSQL 8.1.</i> [PE]: Desinstale PostgreSQL 8.1.
No se traduce correctamente un verbo (4.b.3.1: mistranslation)	99	[TA]: mantenga su ataque HTTP servidor y cierren lo. [Original]: <i>Halten Sie Apache HTTP Server an und schließen Sie ihn.</i> [PE]: Detenga el servidor Apache HTTP y ciérrelo.
No se traduce correctamente una preposición (4.a.1.1: mistranslation)	85	[TA]: * * certificado bajo www.wilo.de / legal [Original]: ** <i>Zertifikat abrufbar unter www.wilo.de/rechtliches.</i> [PE]: * * certificado disponible en www.wilo.de / información legal

Tabla 188: Ejemplos de errores más frecuentes independientemente del motor seleccionado

Los errores de índole sintáctica y la traducción errónea de sustantivos ya aparecen como errores frecuentes en la primera fase del análisis.

También pudimos establecer en que macrocategoría marcamos más errores. En la cuarta macrocategoría, es decir errores que tienen que ver con la microestructura, marcamos más errores, en total 1063; a continuación encontramos 584 errores en la tercera macrocategoría, aquella categoría en la que agrupamos los errores que se producen por las características propias del castellano; en la primera macrocategoría de errores independientes de la combinación lingüística marcamos 96 errores, en la quinta categoría de errores relacionados con el género o redacción del texto original encontramos 48 errores y en la segunda macrocategoría de errores provocados por las características propias del alemán 42 errores. En la sexta macrocategoría marcamos 16 errores, es decir errores que no pudimos ajustar en otras macrocategorías.

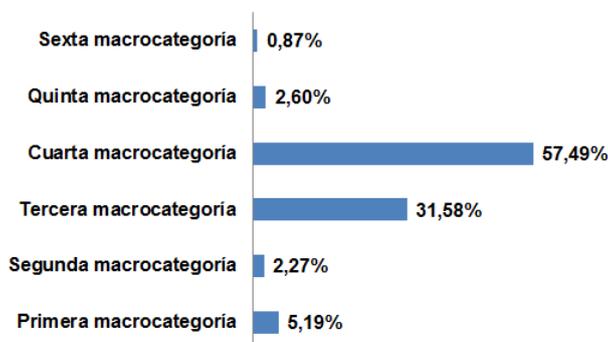


Ilustración 54: porcentaje de errores marcados según las macrocategorías

8. Clasificación de errores para la sistematización de la posesición

En el siguiente capítulo veremos, como también lo hicimos en la primera fase del análisis, qué tipos de errores nos encontramos teniendo en cuenta el motor que se seleccionó. En primer lugar veremos las categorías de errores más frecuentes, pero también veremos la cantidad de errores que se presentaron según la macrocategoría.

8.4.1 Errores encontrados según el motor empleado

Algunos de los resultados vistos como globales se repiten en cada motor por separado, sin embargo se dan en una frecuencia diferente. Como vimos en la primera fase del análisis en las propuestas del primer motor estadístico se interpretaron erróneamente las etiquetas, y en todos las propuestas encontramos errores en la sintaxis. Veamos cómo clasificamos los errores en esta fase según el motor seleccionado.

8.4.2 Errores encontrados en las propuestas del motor SMT1

24 propuestas del primer motor estadístico fueron marcadas como correctas y no tuvieron que ser poseditadas. Este motor ocuparía así el segundo lugar después del RBMT1. En las propuestas escogidas para la posesición marcamos un total de 378 errores, que son de las subcategorías error de sintaxis, traducción errónea de sustantivos, adición de determinantes, omisión de verbos y traducción errónea de preposiciones y problemas de transferencia sintáctica de compuestos nominales. En la gráfica a continuación podemos ver que las primeras cuatro categorías fueron marcadas casi tantas veces y que no hay una categoría que resalte especialmente.

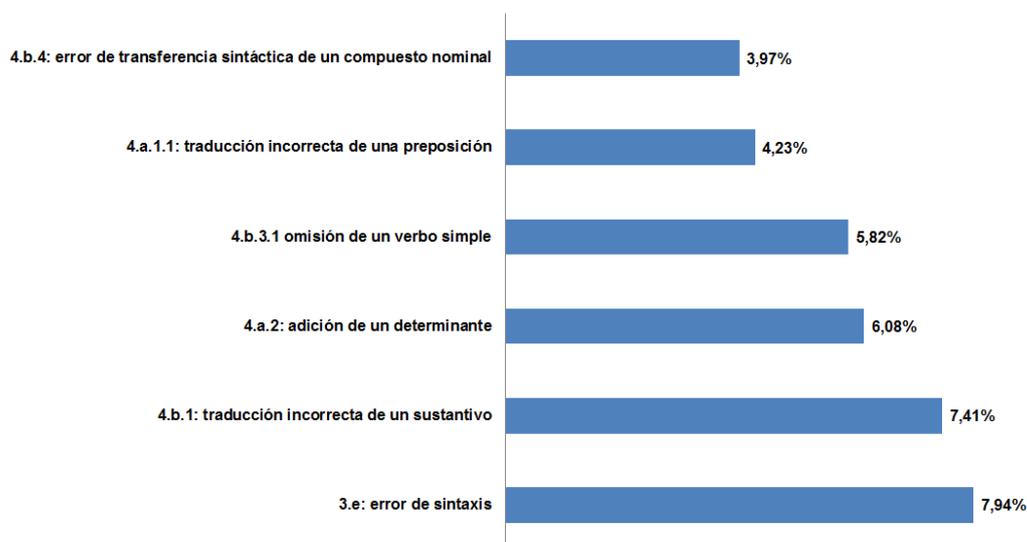


Ilustración 55: Errores más frecuentes en las propuestas de SMT1

8. Clasificación de errores para la sistematización de la posesición

Para poder entender claramente el tipo de errores daremos a continuación ejemplos representativos, como en las tablas anteriores también indicaremos el número de errores marcados:

Categoría de error	Número	Ejemplo
Sintaxis (3.e: misplaced)	30	[TA]: Altamente eficaces herramientas para los planificadores. [Original]: <i>Hocheffiziente Tools für Planer.</i> [PE]: <ut/> Herramientas altamente eficaces para los planificadores.
Traducción errónea de sustantivos (4.b.1: mistranslation)	28	[TA]: Esto se logrará mediante 5 maliges pitidito del? p-v símbolo. [Original]: <i>Dies wird durch 5 maliges Blinken des ?p-v Symbols angezeigt.</i> [PE]: Esto se indicará mediante el parpadeo 5 veces del símbolo ?p-v.
Se añade un determinante (4.a.2: addition)	23	[TA]: Las temperaturas del agua a temperatura ambiente contribuye con un máximo del 40 ° C. [Original]: <i>Wassertemperaturen bei max. Umgebungstemperatur +40 °C.</i> [PE]: Temperaturas del agua a temperatura ambiente máxima de +40 °C
Omisión de verbos (4.b.3: omission)	22	[TA]: Todos los parámetros en varios mapas de registro, para facilitar una ilustración. [Original]: <i>Alle Parameter sind in verschiedenen Registerkarten angeordnet, um eine Veranschaulichung zu vereinfachen.</i> [PE]: Todos los parámetros están en varias fichas de registro, para facilitar su ilustración.
Traducción errónea de preposiciones (4.a.1.1: mistranslation)	16	[TA]: * * certificado bajo www.wilo.de / legal. [Original]: <i>** Zertifikat abrufbar unter www.wilo.de/rechtliches.</i> [PE]: * * certificado disponible en www.wilo.de / información legal
Error de transferencia sintáctica de un compuesto nominal (2.c.1: mistranslation)	15	[TA]: En un doble clic o a los aparatos nombre tendrá acceso a su parámetros frase: [Original]: <i>Bei einem Doppelklick auf den Gerätenamen erhalten Sie Zugang zu ihrem Parametersatz:</i> [PE]: Haciendo doble clic en el nombre de los aparatos tendrá acceso a su conjunto de parámetros:

Tabla 189: Ejemplos de errores más frecuentes en las propuestas del smt1

8. Clasificación de errores para la sistematización de la posesición

En relación con las macrocategorías nos encontramos con la siguiente distribución. De los 378 errores marcados la mayoría pertenecen a la fue la cuarta macrocategoría (229 errores) y a continuación nos encontramos con la tercera macrocategoría (104 errores). Un número menos considerable de errores se marcaron en las macrocategorías restantes: 19 errores en la primera, 14 en la quinta, 7 en la segunda y 5 en la sexta macrocategoría. En la siguiente gráfica ilustramos el porcentaje de errores marcados por macrocategoría.

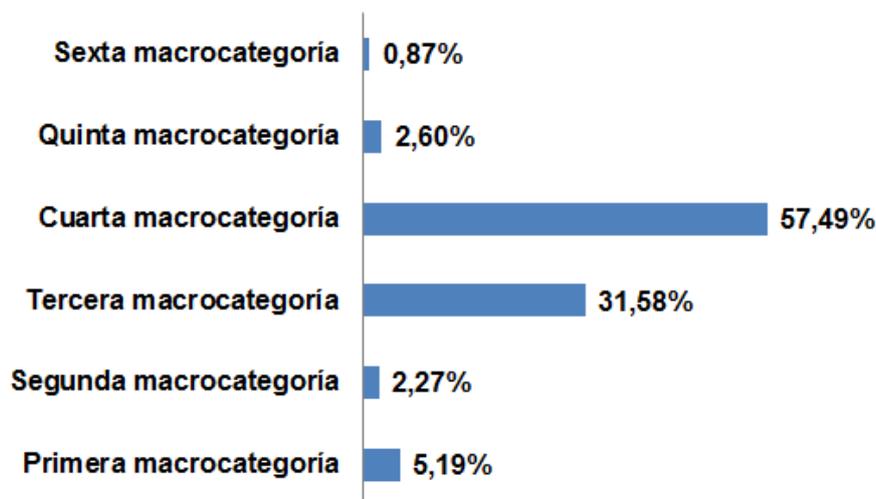


Ilustración 56: Porcentaje de errores encontrados en las propuestas del smt1 según las macrocategorías definidas

8.4.3 Errores encontrados en las propuestas del motor SMT2

Solo 14 propuestas del segundo motor estadístico fueron marcadas como correctas y no tuvieron que ser poseídas. En las propuestas escogidas para la posesición marcamos un total de 275 errores, que son de las subcategorías adición de determinantes, errores de flexión nominal de número, errores de sintaxis, tratamiento erróneo de las etiquetas que se encontraban en el original y traducción también errores de transferencia léxica y sintáctica de compuestos nominales y/o composiciones. En la gráfica a continuación podemos ver que no hay diferencia muy grande en el número de errores marcados. También aparece por primera vez en esta fase el problema del tratamiento incorrecto de las etiquetas, así como los errores que se dan en castellano cuando no se ajusta correctamente el número en un sintagma.

8. Clasificación de errores para la sistematización de la posesición

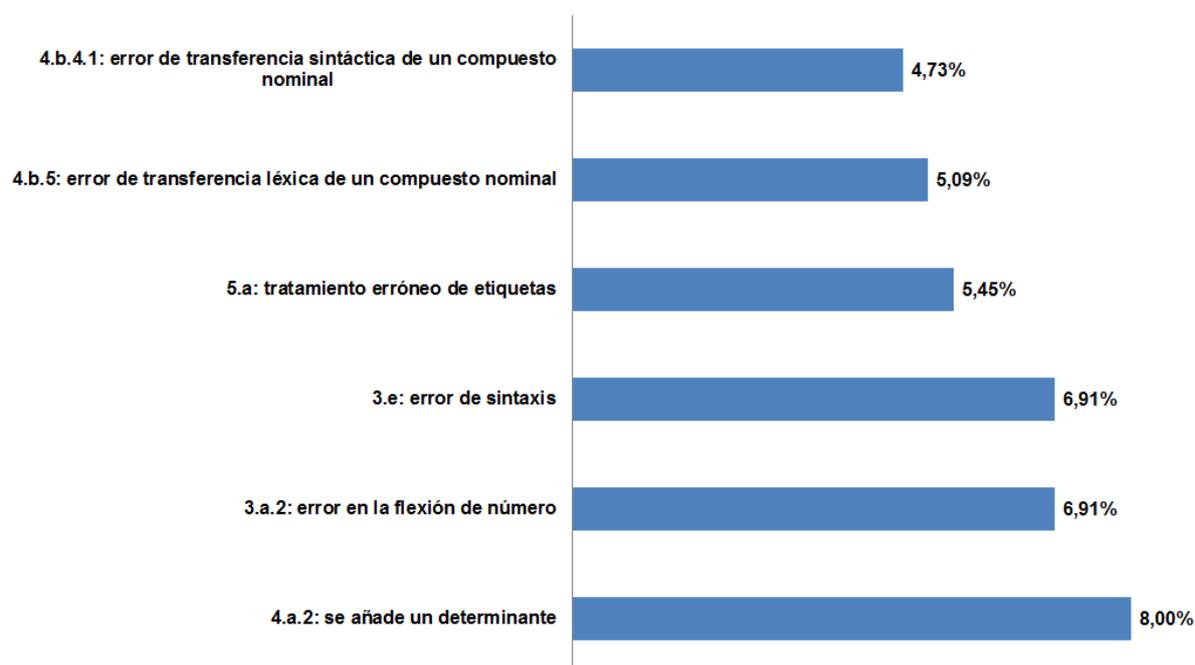


Ilustración 57: Errores más frecuentes en las propuestas de SMT2

Para poder entender claramente a qué tipo de errores nos referimos, daremos a continuación ejemplos representativos. En algunos casos se presenta más de un error y por ello los marcamos en color. También, como en el caso de la traducción errónea de los verbos, se presentan dos errores en la misma palabra, no solo no se traduce el equivalente correcto del verbo en el caso concreto, sino que además la conjugación es errónea.

Categoría de error	Número	Ejemplo
Se añade un determinante (4.a.2: addition)	22	[TA]: La duración de bombeo [Original]: <i>Pump down Dauer:</i> [PE]: Duración de bombeo de vacío:
Errores en la flexión de número (3.a.2: mistranslation)	19	[TA]: Los usuarios pueden sufrir lesiones graves ().. [Original]: <i>Der Benutzer kann (schwere) Verletzungen erleiden.</i> [PE]: El usuario puede sufrir lesiones (graves)..
Errores de sintaxis (3.e: misplaced)	19	[TA]: Las advertencias de la administración . [Original]: <i>Verwaltung der Warnhinweise.</i> [PE]: Administración de las advertencias.

8. Clasificación de errores para la sistematización de la posesición

Categoría de error	Número	Ejemplo
Tratamiento erróneo de etiquetas (5.a: questionable)	15	[TA]: Suceda Frm ID = "17". / suceda. B: Reducción de la noche suceda /b [Original]: <i><Frm id="17"/> Nachtabsenkung:</i> [PE]: <i><Frm id="17"/> Reducción nocturna:</i>
Errores de transferencia léxica de compuestos nominales (4.b.4: mistranslation)	14	[TA]: En la modalidad de reposo bomba el botón rojo durante 10 segundos. [Original]: <i>Im Ruhemodus der Pumpe den roten Knopf 10 Sekunden lang drücken.</i> [PE]: En el modo de reposo de la bomba pulse el botón rojo durante 10 segundos.
Errores de transferencia sintáctica de compuestos nominales (2.c.1: mistranslation)	13	[TA]: La eficiencia y altamente rentable, inmediatamente para el medio ambiente y las empresas. [Original]: <i>Und Hocheffizienz rechnet sich sofort - für die Umwelt und die Unternehmen.</i> [PE]: La alta eficiencia es rentable inmediatamente para el medio ambiente y las empresas.

Tabla 190: Ejemplos de errores más frecuentes en las propuestas del smt2

En relación con las macrocategorías nos encontramos con una distribución similar a la que se dio en las propuestas del smt1. De los 275 errores marcados la mayoría pertenecen a la también a la cuarta macrocategoría (144 errores) y a continuación nos encontramos con la mayor cantidad de errores en la tercera macrocategoría (91 errores). Un número menos considerable de errores se marcaron en las macrocategorías restantes: igual que en las propuestas del smt marcamos 19 errores en la primera macrocategoría, 15 en la quinta, 4 en la sexta y 2 en la segunda macrocategoría. En la siguiente gráfica ilustramos el porcentaje de errores marcados por macrocategoría.

8. Clasificación de errores para la sistematización de la posesición

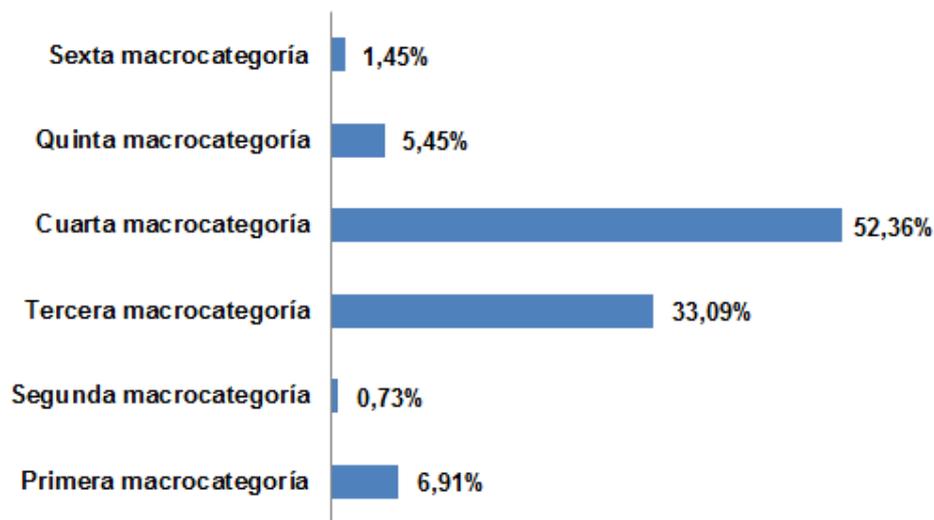


Ilustración 58: Porcentaje de errores encontrados en las propuestas del smt2 según las macrocategorías definidas

8.4.4 Errores encontrados en las propuestas del motor RBMT1

Como enunciamos al comienzo la mayor parte de las propuestas seleccionadas para la posesición eran propuestas de este motor de traducción; además la mayoría de las propuestas marcadas como correctas también fueron las realizadas con este motor (101). A diferencia de los motores estadísticos el mayor número de errores no se dieron en la sintaxis, sino en la traducción errónea de sustantivos. Los siguientes problemas tienen que ver con la microestructura, en concreto, con la traducción errónea de verbos y preposiciones. Y finalmente se presentaron errores en la transferencia sintáctica de compuestos nominales. Aquí el número de errores marcado es mayor, porque también fueron muchas más las propuestas seleccionadas para la posesición en comparación con los otros motores.

8. Clasificación de errores para la sistematización de la posesición

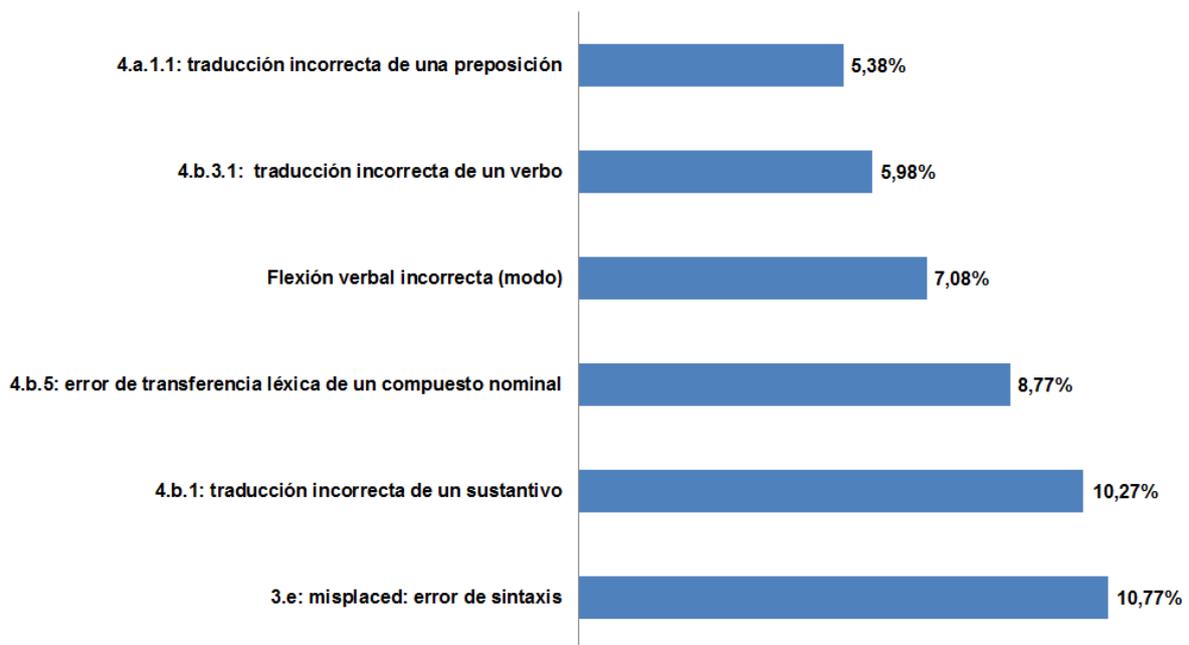


Ilustración 59: Errores más frecuentes en las propuestas de RBMT1

Para poder entender claramente a qué tipo de errores nos referimos, daremos a continuación ejemplos representativos. Como en los casos anteriores es posible que en un ejemplo se encuentren errores de diferentes categorías, por ello marcamos en color el error a que nos referimos en cada caso. En el último caso, por ejemplo, se presentan dos errores en la traducción de compuestos nominales, pero solo marcamos uno de ellos.

Categoría de error	Número	Ejemplo
Errores de sintaxis (3.e: misplaced)	108	[TA]: A continuación hace sobre añadir clic, con eso está registrado el aparato. [Original]: <i>Anschließend klicken Sie auf Hinzufügen, damit ist das Gerät registriert.</i> [PE]: A continuación haga clic en "agregar", para que quede registrado el aparato.
Traducción errónea de sustantivos (4.b.1: mistranslation)	103	[TA]: Después se indican dos otras noticias: [Original]: <i>Daraufhin werden zwei weitere Nachrichten angezeigt:</i> [PE]: Después se indican otros dos mensajes:

8. Clasificación de errores para la sistematización de la posesición

Categoría de error	Número	Ejemplo
Errores de transferencia léxica de compuestos nominales (4.b.4: mistranslation)	88	[TA]: Supresión de la altura de fomentar en m. [Original]: <i>Einstellung der Förderhöhe in m.</i> [PE]: Ajuste de la altura de impulsión en m.
Flexión verbal incorrecta (modo) (3.b.2: mistranslation)	71	[TA]: Automáticamente actualizan [Original]: <i>Automatisch aktualisieren</i> [PE]: Actualizar automáticamente
Traducción errónea de verbos (4.b.3: mistranslation)	60	[TA]: Haga clic en Abandono , para cerrar el programa de instalación. [Original]: <i>Klicken Sie auf Verlassen, um das Installationsprogramm zu schließen.</i> [PE]: Haga clic en Salir para cerrar el programa de instalación.
Traducción errónea de preposiciones (4.a.1: mistranslation)	54	[TA]: $6 =$ altura de fomentar máxima en m en $Q = 0 \text{ m}^3/\text{h}$ [Original]: $6 =$ maximale Förderhöhe in m bei $Q = 0 \text{ m}^3/\text{h}$ [PE]: $6 =$ altura de impulsión máxima en m para $Q = 0 \text{ m}^3/\text{h}$

Tabla 191: Ejemplos de errores más frecuentes en las propuestas del rbmt1

En lo que tiene que ver con las macrocategorías nos encontramos con una distribución no tan diferente de la que se dio en las propuestas de los motores estadísticos. Aquí también se marcaron la mayoría de los 1003 errores en la cuarta macrocategoría (602 errores) y a continuación nos encontramos la mayor cantidad de errores en la tercera macrocategoría (310 errores). Un número menos considerable de errores se marcaron en las macrocategorías restantes: 43 errores en la primera macrocategoría, 28 en la segunda, 14 en la quinta y 6 errores en la sexta macrocategoría. En la siguiente gráfica ilustramos el porcentaje de errores marcados por macrocategoría.

8. Clasificación de errores para la sistematización de la posesición

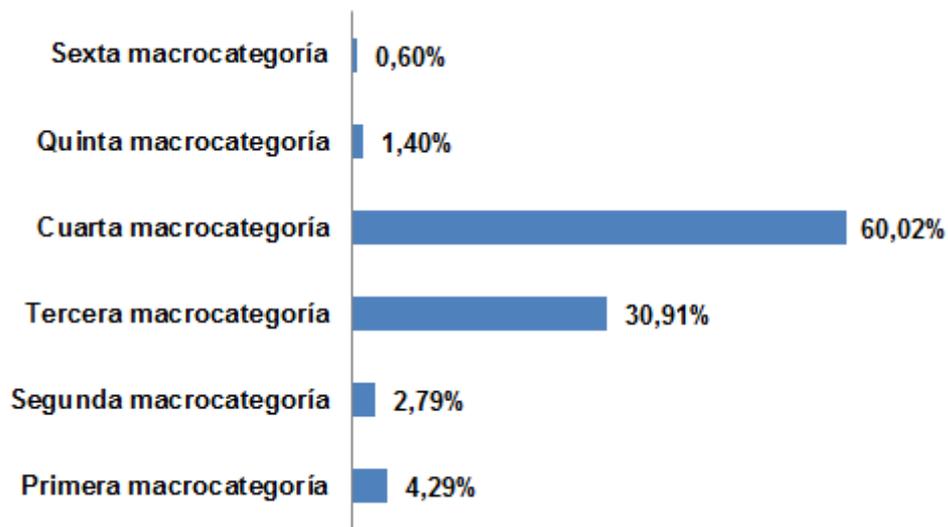


Ilustración 60: Porcentaje de errores encontrados en las propuestas del rbmt1 según las macrocategorías definidas

8.4.5 Errores encontrados en las propuestas del motor RBMT2

En las propuestas que se realizaron con este tipo de motor marcamos un total de 193 errores. Aquí volvemos a encontrar en primer lugar problemas de orden sintáctico. Los otros tipos de problemas se dan principalmente en la microestructura, es decir, en la traducción errónea de palabras con y sin carga semántica, en la omisión de preposiciones. También encontramos en los errores más frecuentes problemas de flexión verbal.

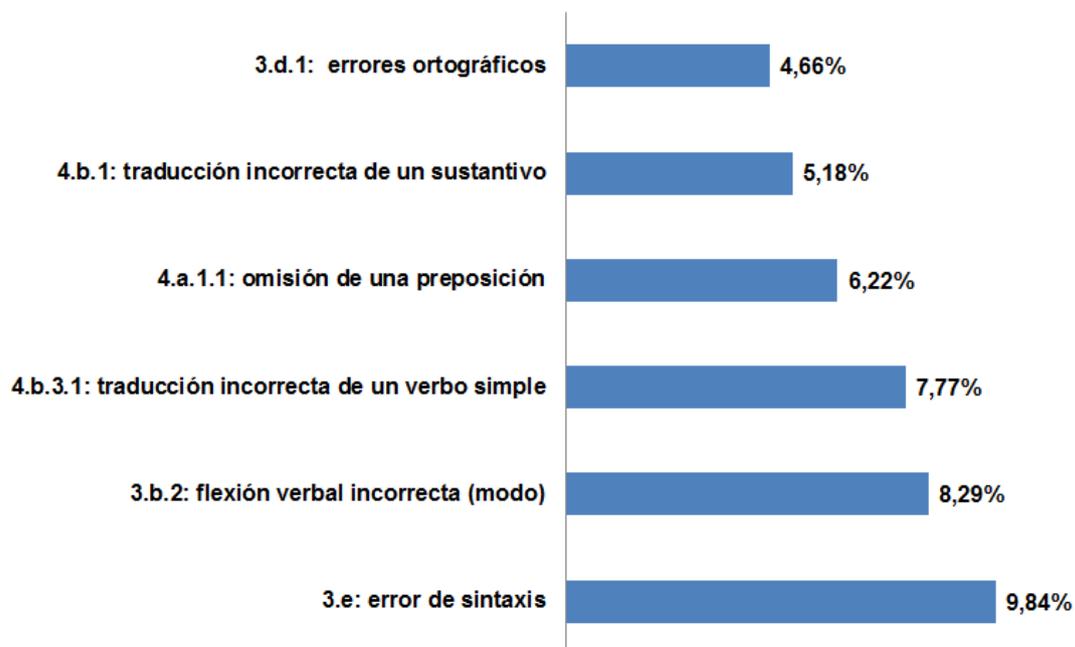


Ilustración 61: Errores más frecuentes en las propuestas de RBMT2

8. Clasificación de errores para la sistematización de la posesición

A continuación algunos ejemplos de los errores más frecuentes. Como en los casos anteriores en algunos de los ejemplos se presentan varios errores y solo marcaremos el error de la categoría que queremos ilustrar:

Categoría de error	Número	Ejemplo
Error de sintaxis (3.e: misplaced)	19	[TA]: El paso próximo en la Wizard por es señalado clic en las Schaltfläche de Lejanas. [Original]: Durch Klicken auf die Schaltfläche Weiter wird der nächste Schritt im Wizard angezeigt. [PE]: Haciendo clic en el botón "siguiente" se indicará el siguiente paso en el asistente.
Flexión verbal incorrecta (modo) (3.b.2: mistranslation)	16	[TA]: Desinstalar PostgreSQL 8.1. [Original]: Deinstallieren Sie PostgreSQL 8.1. [PE]: Desinstale PostgreSQL 8.1.
Traducción errónea de verbos (4.b.3.1: mistranslation)	15	[[TA]: Fallan de funciones importantes del producto, la planta/, [Original]: Versagen wichtiger Funktionen des Produktes/der Anlage, [PE]: Fallo de funciones importantes del producto / planta,
Se omite una preposición (4.a.1: omission)	12	[TA]: 512 MB memoria de trabajo (RAM) [Original]: 512 MB Arbeitsspeicher (RAM) [PE]: 512 MB de memoria de trabajo (RAM)
No se traduce correctamente un sustantivo (4.b.1: mistranslation)	10	[TA]: Aquellos posiblemente al amparo de es señalado del suceso de setup noticia siguiente. [Original]: Möglicherweise wird im Rahmen des Setupvorgangs die folgende Nachricht angezeigt. [PE]: Es posible que durante el proceso de instalación se muestre el siguiente mensaje:
Errores ortográficos (3.d.1: questionable)	9	[TA]: Política, sociedad y industria, son de acuerdo se: [Original]: Politik, Gesellschaft und Industrie sind sich einig: [PE]: Política, sociedad e industria están de acuerdo en que

Tabla 192: Ejemplos de errores más frecuentes en las propuestas del rbmt2

Con respecto a las macrocategorías la distribución de los errores es igual que la de los errores encontrados en las propuestas del rbmt1, solo difieren en la cantidad de errores marca-

8. Clasificación de errores para la sistematización de la posesición

dos por categoría. Aquí la mayoría de los 193 errores macados en total, la mayoría se encuentran en la cuarta macrocategoría (88 errores) y a continuación nos encontramos la mayor cantidad de errores en la tercera macrocategoría (79 errores). Un número menor de errores se marcaron en las macrocategorías restantes: 15 errores en la primera macrocategoría, 5 en la segunda y en la quinta y solo 1 error en la sexta macrocategoría. En la siguiente gráfica ilustramos el porcentaje de errores marcados por macrocategoría.

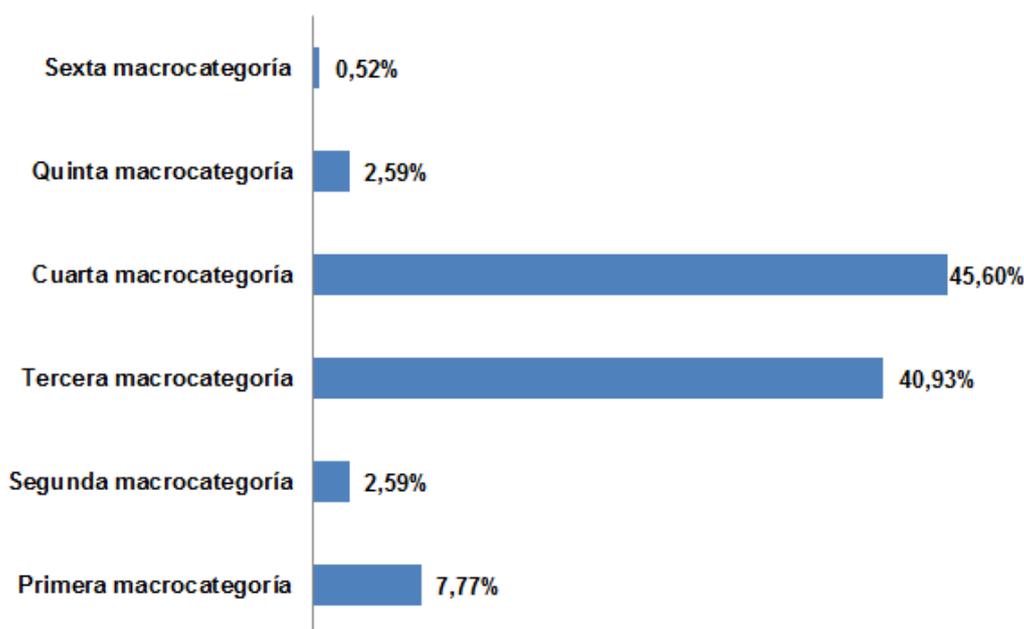


Ilustración 62: Porcentaje de errores encontrados en las propuestas del rbmt2 según las macrocategorías definidas

Encontramos además algunos casos en los que el error no tenía una explicación posible generalmente porque se añadieron palabras que no estaban en el original. En la siguiente tabla se encuentran estos casos, en la tabla solo marcamos el error que no tenía justificación y no tuvimos en cuenta los otros errores que se presentaban en el segmento en cuestión:

Segmento	Descripción
<p>[TA]: A partir de ahora el sistema instala los unos a los otros productos diferentes que están hechos en AKO 5004 participar.</p> <p>[Original]: <i>Ab jetzt installiert das System nacheinander verschiedene Produkte, die an AKO 5004 beteiligt sind.</i></p>	<p>se añaden la forma verbal 'están hechos', que no se encuentra en el original ni hace falta en castellano</p>

8. Clasificación de errores para la sistematización de la posesición

Segmento	Descripción
[PE]: A partir de ahora el sistema instala uno tras otro diferentes productos que intervienen en AKO 5004.	
<p>[TA]: Suceda ut / me Deinstallieren usted ataque HTTP servidor 2.2.3. ut / me suceda ut / me para la Deinstallation importante que este producto la Windows-Funktion suceda ut / me suceda ut / me programas o eliminar las cosas ut / me suceda ut / me (sistema de lanzamiento me gestión me software).</p> <p>[Original]: <i><ut/>Deinstallieren Sie Apache HTTP Server 2.2.3. <ut/> <ut/>Für die Deinstallation dieses Produkts müssen Sie die Windows-Funktion <ut/> <ut/>Programme ändern oder entfernen <ut/> <ut/>(Start > Systemsteuerung > Software) verwenden.</i></p> <p>[PE]: <i><ut/>Deinstallieren Sie Apache HTTP Server 2.2.3. <ut/> <ut/>Für die Deinstallation dieses Produkts müssen Sie die Windows-Funktion <ut/> <ut/>Programme ändern oder entfernen <ut/> <ut/>(Start > Systemsteuerung > Software) verwenden.</i></p>	se añade el adjetivo 'importante'
<p>[TA]: <i><ut/>Cuando Usted en la superficie de conectar <ut/> <ut/>Repeler<ut/> <ut/> hacen clic, si el proceso de instalación se termina enseguida.<ut/></i></p> <p>[Original]: <i><ut/>Wenn Sie auf die Schaltfläche <ut/> <ut/>Abweisen<ut/> <ut/> klicken, wird der Installationsvorgang sofort beendet.<ut/></i></p> <p>[PE]: <i><ut/>Cuando usted hace clic en el botón <ut/> <ut/>Cancelar<ut/> <ut/>, el proceso de instalación se termina enseguida.<ut/></i></p>	se añade la conjunción 'si'
<p>[TA]: 1-Puerto puertos USB libres</p> <p>[Original]: <i>1 freier USB-Port</i></p>	se añade el sustantivo 'puertos'

8. Clasificación de errores para la sistematización de la posesición

Segmento	Descripción
[PE]: 1 puerto USB libre	
[TA]: Esta aplicación nos ayudará a desinstalar el producto. [Original]: <i>Diese Anwendung führt Sie durch den Deinstallationsvorgang für das Produkt.</i> [PE]: Esta aplicación le guiará por el proceso de desinstalación del producto.	se añade un guion
[TA]: Suceda. / Tab anual promedio del consumo de electricidad suceda. / NewLine sólo 46,5 kwh del sur de ITV certificados * * [Original]: <i>Durchschnittlicher Jahresstromverbrauch von nur 46,5 kWh vom TÜV SÜD zertifiziert**</i> [PE]: <Tab/>Promedio del consumo anual de electricidad <NewLine/>sólo 46,5 kwh certificado por TÜV SÜD * *	no se traduce correctamente el nombre propio 'TÜV SÜD' (sur de ITV / correcta: TÜV SÜD)
[TA]: Área de control hasta tres veces mayor que en bombas convencionales reguladas electrónicamente [Original]: <i>Regelbereich bis zu dreimal größer als bei herkömmlichen elektronisch geregelten Pumpen</i> [PE]: Área de control hasta tres veces mayor que en bombas convencionales reguladas electrónicamente	error tipográfico

Tabla 193: Segmentos en que encontramos errores sin justificación

Aunque el foco principal de este segundo análisis no es corregir la posesición encontramos 10 casos en los que pudimos detectar errores. Decidimos mencionarlos porque pueden ayudarnos a definir reglas para la posesición en nuestra combinación lingüística.

8. Clasificación de errores para la sistematización de la posesición

Segmento	Descripción
<p>[TA]: (sección<field name="\$tbsheetnum"/> de<field name="\$tbsheetcount"/>)</p> <p>[Original]: (<i>Abschnitt <field name="\$tbsheetnum"/> von <field name="\$tbsheetcount"/></i>)</p> <p>[PE]: (sección <field name="\$tbsheetnum"/> de<field name="\$tbsheetcount"/>)</p>	Falta un espacio en blanco después de la preposición 'de'.
<p>[TA]: * basada en la combinación energética de la UE</p> <p>[Original]: * <i>Basierend auf EU-Energiemix.</i></p> <p>[PE]: * Basada en la combinación energética de la UE,</p>	Error ortográfico (se añade una coma).
<p>[TA]: <ut/>Para aceptar el acuerdo de licencia y su contenido, deben Usted el campo de opción <ut/> <ut/>He leído las condiciones contractuales y acepto éstos<ut/> <ut/> seleccionan y a continuación en la superficie de conectar <ut/> <ut/>Aceptar<ut/> <ut/> hacer clic.</p> <p>[Original]: <ut/>Um die Lizenzvereinbarung und ihren Inhalt zu akzeptieren, müssen Sie das Optionsfeld <ut/> <ut/>Ich habe die Vertragsbedingungen gelesen und nehme diese an<ut/> <ut/> auswählen und anschließend auf die Schaltfläche <ut/> <ut/>Annehmen<ut/> <ut/> klicken.</p> <p>[PE]: <ut/>Para aceptar el acuerdo de licencia y su contenido, deben usted seleccionar el campo de opción <ut/> <ut/>He leído las condiciones contractuales y las acepto<ut/> <ut/> y a continuación hacer clic en el botón <ut/> <ut/>Aceptar<ut/> <ut/>.</p>	No se conjuga correctamente el verbo 'deber' (persona).
<p>[TA]: Paquetes Adapt es una adaptación del dinámica debe valor en la parte de la carga de la bomba, en menos de la mitad del volumen de interpretación de la electricidad.</p>	Se trata de traducir erróneamente el nombre propio 'Dynamic Adapt'.

8. Clasificación de errores para la sistematización de la posesición

Segmento	Descripción
<p>[Original]: <i>Dynamic Adapt ist eine dynamische Anpassung des Sollwertes im Teillastbereich der Pumpe, bei weniger als dem halben Auslegungsvolumenstrom.</i></p> <p>[PE]: Dinamic Adapt es una adaptación dinámica del valor de consigna en el modo de carga parcial de la bomba, a menos de la mitad del flujo volumétrico de referencia.</p>	
<p>[TA]: Confirmar supresión por presión del botón rojo.</p> <p>[Original]: <i>Einstellung durch Drücken des roten Knopfes bestätigen.</i></p> <p>[PE]: Confirmar el ajuste por presión del botón rojo.</p>	<p>No se corrige el error de la traducción de la expresión preposicional 'durch Drücken'.</p>
<p>[TA]: Consiste el peligro, dañar el producto, la planta/.</p> <p>[Original]: <i>Es besteht die Gefahr, das Produkt/die Anlage zu beschädigen.</i></p> <p>[PE]: Exsiste el peligro de dañar el producto / la planta.</p>	<p>Errores ortográficos: no se escribe correctamente el verbo y se añaden espacios antes y después de la barra.</p>
<p>[TA]: Con el Wilo-Connector en muchos productos sin herramienta, y gracias a tecnología de botón de rojos y visualizador funciona también la puesta en marcha fácilmente y rápido</p> <p>[Original]: <i>Mit dem Wilo-Connector bei vielen Produkten ohne Werkzeug, und dank Rote-Knopf-Technologie und Display geht auch die Inbetriebnahme kinderleicht und blitzschnel.</i></p> <p>[PE]: Con el Conector Wilo funciona también la puesta en marcha fácil y rápida en muchos productos, sin herramientas y gracias a la tecnología de botón rojo y visualizador</p>	<p>Error de sintaxis.</p>
<p>[TA]: Si el sistema indica la lista de los aparatos encontrados, hace sobre eso clic para registrar éstos en el sistema.</p>	<p>Se añade una de las preposiciones</p>

8. Clasificación de errores para la sistematización de la posesición

Segmento	Descripción
<p>[Original]: <i>Wenn das System die Liste der gefundenen Geräte anzeigt, klicken Sie darauf, um diese im System zu registrieren.</i></p> <p>[PE]: Si el sistema indica la lista de los equipos encontrados, haga clic en sobre ella para registrar éstos en el sistema.</p>	
<p>[TA]: Si cliquea sobre las observaciones de advertir de tarjeta de registro, recibe lo siguiente:</p> <p>[Original]: <i>Wenn Sie die Registerkarte Warnhinweise anklicken, erhalten Sie Folgendes:</i></p> <p>[PE]: Si hace clic sobre las advertencias de la tarjeta de registro, recibe la siguiente indicación:</p>	<p>El modo verbal seleccionado no es correcto, se debe utilizar el subjuntivo.</p>
<p>[TA]: Para la creación de un usuario nuevo tiene que hacer sólo sobre la superficie de conectar correspondiente que conduce al anuncio de la pantalla siguiente clic:</p> <p>[Original]: <i>Zur Erstellung eines neuen Anwenders müssen Sie nur auf die entsprechende Schaltfläche klicken, die zur Anzeige des folgenden Bildschirms führt:</i></p> <p>[PE]: Para la creación de un usuario nuevo sólo tiene que hacer clic sobre la superficie de conexión correspondiente que conduce al anuncio de la pantalla siguiente:</p>	<p>No se traduce correctamente el sustantivo 'Schaltfläche' (botón)</p>

Tabla 194: Errores detectados en la posesición

8.5 Conclusiones del segundo análisis

En esta fase del análisis pudimos comprobar que el motor más efectivo, es decir, el que fue preferido por los poseedores, es el rmbt1 y el menos favorecido el rbmt2.

8. Clasificación de errores para la sistematización de la posesición

El problema del tratamiento incorrecto de las etiquetas se presentó básicamente en los motores basados en estadísticas. En este segundo análisis no marcamos las palabras que se añadieron en los casos en que se trataron las etiquetas como texto, ya que estas adiciones se dieron por dicho tratamiento erróneo. Tampoco marcamos más los errores que denominamos en el primer análisis como errores pragmáticos, es decir aquellos casos en que no se cumplía la función comunicativa del original. Esto debido a que los segmentos que se escogieron para la posesición generalmente transmitían el mensaje del texto original y en estos segmentos solo encontramos errores que dificultaban la fluidez en el idioma meta. Encontramos además que los errores de sintaxis, la traducción de sustantivos y de compuestos nominales son los errores más frecuentes a nivel general.

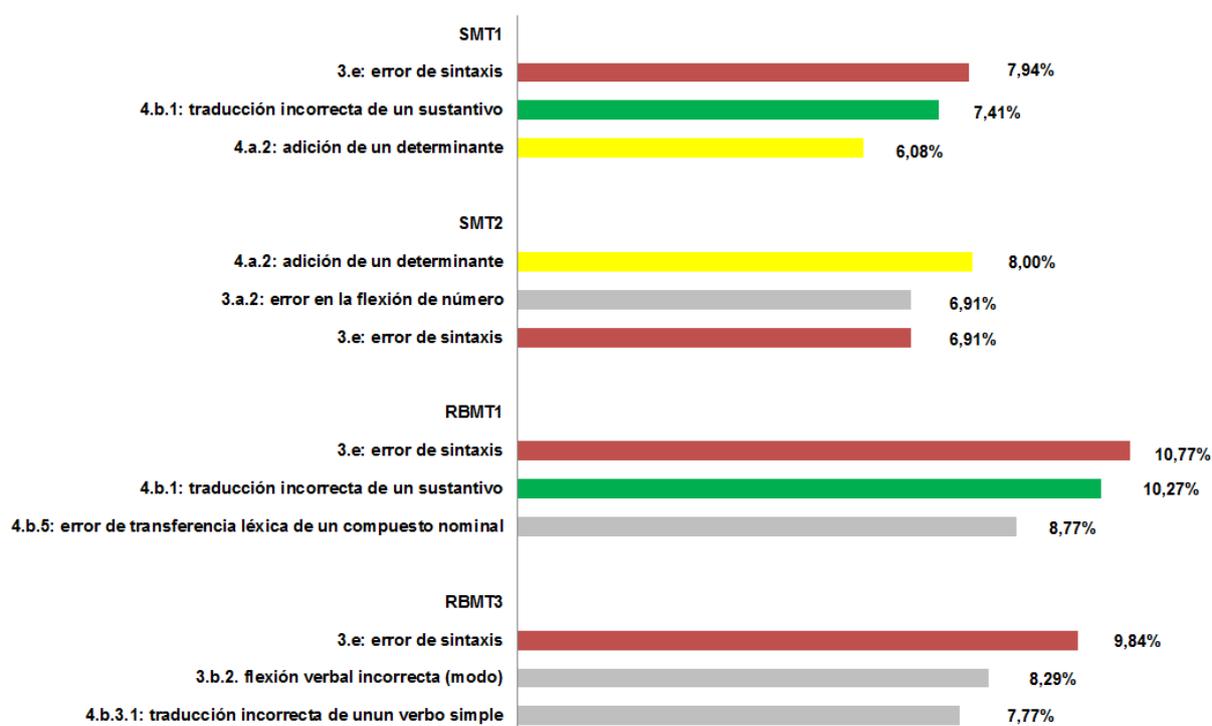


Ilustración 63: Comparación de los errores más frecuentes según el motor seleccionado

Adicionalmente pudimos constatar que ciertas combinaciones de adverbios o preposiciones son problemáticas para todos los motores y del mismo modo que ciertos errores repetitivos nos pueden ayudar a formular reglas para la posesición en nuestra combinación lingüística. En ciertos casos nos encontramos con la dificultad de clasificar un error en una categoría determinada, ya que dicho error podría ajustarse a varias categorías. Esto es el caso, por ejemplo, de los errores en la traducción de composiciones con un anglicismo, que podían con-

8. Clasificación de errores para la sistematización de la posesición

tarse como errores de la categoría 1.a.1 (presencia de léxico extranjero en el original) o errores de la categoría 2.d (errores en la transferencia sintáctica de composiciones) e incluso en la categoría 4.b.4 (errores en la transferencia léxica de compuestos nominales y composiciones). En estos casos y otros similares marcamos el error que consideramos más representativo, pero ello no excluye el que pueda ser considerado de otra manera.

PARTE III: RESULTADOS

9. Incidencia de cada categoría de errores en el corpus de análisis

Después de haber presentado en el capítulo 7 el análisis preliminar y en el capítulo 8 nuestra clasificación de errores para la sistematización de la posesición, en este capítulo analizaremos qué resultados hemos obtenido con dicha clasificación. Aquí vamos a ver en primer lugar, qué incidencia tuvo cada macrocategoría y cómo se distribuyen los errores. En segundo lugar vamos a analizar la incidencia de los errores teniendo en cuenta el tipo de motor seleccionado y los motores individualmente en relación con las macrocategorías y categorías que planteamos anteriormente. Es decir, evaluaremos la incidencia por tipo de motor y por motor. Nos interesa ver si se da alguna regularidad entre los errores que hemos identificado y algún tipo de motor concreto. Por otro lado, también observaremos cada tipo de motor en relación con cada categoría y cada macrocategoría. Y cada motor concreto por cada categoría. De este modo comprobaremos si los errores que los poseedores prefirieron editar se pueden vincular con algún tipo de motor concreto o si, por el contrario, su distribución es aleatoria. También comprobaremos si, observando los datos de todos los segmentos con relación a los tipos de motores, es posible deducir estrategias útiles para la tarea de los poseedores.

Para ello vamos a abordar qué patrones o qué repeticiones se observan en primer lugar, desde el punto de vista de la macrocategoría y, en segundo lugar, teniendo en cuenta el motor. En primer término observaremos qué errores se dan por macrocategorías y, después, como se relacionan estos errores con el motor empleado. Luego abordaremos los resultados partiendo de cuál es el origen de cada segmento, para ver si se da algún tipo de predilección sistemática o de error sistemático en la elección. En tercer lugar, analizaremos detalladamente las categorías de errores de mayor incidencia, y especialmente los valores en los que se concretaban esos errores. A partir de los datos obtenidos presentaremos finalmente nuestras sugerencias, en qué punto nuestro análisis nos puede ayudar a sistematizar la formación de traductores para realizar tareas de posesición en la combinación lingüística alemán-castellano; ayudas en la toma de decisiones, qué tipo de formación previa se necesita, si es o no importante revisar el original.

9.1 *Incidencias por macrocategorías*

En este apartado presentamos la distribución de los errores editados tomando como criterio principal las macrocategorías y categorías establecidas en el capítulo anterior. A con-

9. Incidencia de cada categoría de errores en el corpus de análisis

tinuación, este acercamiento también tendrá en cuenta el motor origen de cada segmento editado. Para facilitar el seguimiento de los resultados presentamos en la tabla a los aspectos a que se refiere cada una de las macrocategorías que definimos en el capítulo 8.

Macro categoría	Descripción
Primera	En esta macrocategoría agrupamos todos los errores que se dan, independientemente de la combinación lingüística. Estos errores pueden ser de naturaleza lingüística o extralingüística.
Segunda	Aquí marcamos todos los errores provocados por las características propias del alemán: verbos separables, conjunciones, colocaciones.
Tercera	Bajo esta categoría se encuentran los problemas ocasionados por características propias del castellano, como errores de flexión, concordancia, ortotipografía, estilo y sintaxis.
Cuarta	En esta categoría se agrupan todos los errores en la microestructura, en las palabras sin y con carga semántica.
Quinta	En la quinta categoría incluimos todos los errores que se dieron por problemas o errores en el original.
Sexta	En la sexta categoría incluimos los casos que no pudimos clasificar en ninguna de las categorías anteriores.

Tabla 195: descripción general de las macrocategorías empleadas en el análisis

Gracias a la tabla de conteo pudimos establecer cuántos errores editaron los poseedores (1849 errores marcados) e hicimos dicho cálculo teniendo en cuenta las categorías y macrocategorías que describimos en el capítulo anterior. En la ilustración 58 a continuación, en la que visualizamos el porcentaje de errores marcados por macrocategoría, podemos ver que en la cuarta macrocategoría encontramos la mayor cantidad de ediciones catalogadas. De los 1834 errores detectados 1063 errores se dieron a nivel de la microestructura. En segundo lugar, marcamos 584 ediciones provocados por las características propias del castellano, como la flexión verbal o concordancia de género o número. La tercera macrocategoría en la que marcamos más errores es la primera, aquella en que anotamos los errores que se dan independientemente de la combinación lingüística. Aquí encontramos en total 96 errores. Casi hallamos el mismo número de ediciones en la quinta macrocategoría, errores relacionados con el género del texto o la redacción del original (48 errores en total), y en la segunda macrocategoría errores provocados por las características propias del alemán (42 errores en total).

9. Incidencia de cada categoría de errores en el corpus de análisis

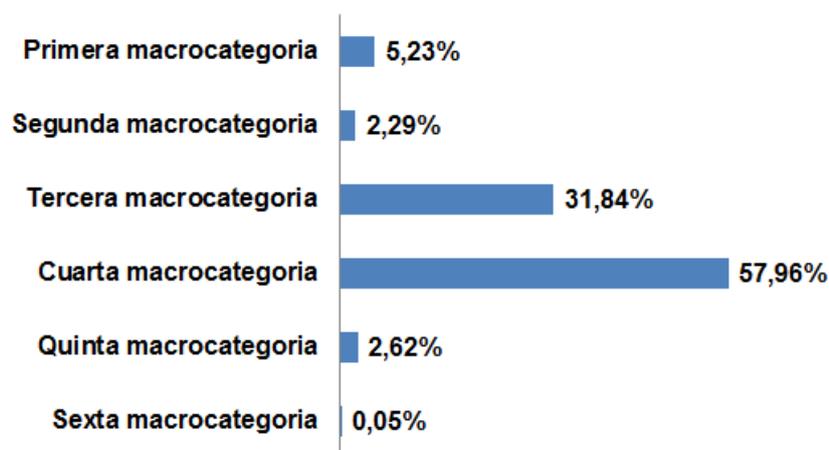


Ilustración 64: Porcentaje de errores encontrados según las macrocategorías

Tal como queda patente, el mayor número de ediciones se realizaron en las macrocategorías cuarta y tercera; esto nos revela que los problemas principales que encuentran los poseedores en nuestra combinación lingüística tienen que ver más con las características propias del castellano, y no tanto con las peculiaridades del alemán. Dado que se trata de los segmentos que los poseedores eligieron utilizar como punto de partida para su traducción final, también cabe concluir que prefieren editar este tipo de errores que otros errores que puedan requerir un nivel de edición mayor o más complejo. Pero para poder darnos una idea más clara de los errores que encontraron y editaron los poseedores analizaremos cada macrocategoría por separado detalladamente.

9.1.1 Incidencia de los errores en la primera macrocategoría

En la ilustración a continuación podemos ver qué errores se presentaron en la primera macrocategoría. La mayor cantidad de problemas se encontraron en el tratamiento de nombres propios, abreviaturas o siglas; ya sea bien porque no se tradujeron correctamente —en 32 casos— o porque se tradujeron en casos en que no era necesario hacerlo (21 errores). También es fuente de errores la presencia de léxico extranjero, que como en el caso de los nombres propios o bien se traducen erróneamente (12 errores) o no se traducen (13 errores). Además, encontramos errores de funcionalidad de elementos tipográficos, que o se colocaron en la posición errónea (7 casos), se adicionaron innecesariamente (también 7 errores) o se omitieron (2 errores). Encontramos 2 errores en la traducción de hipertexto (URLs). A pesar de que en nuestra tipología había descrito como una categoría de esta macrocategoría los errores rela-

9. Incidencia de cada categoría de errores en el corpus de análisis

cionados con convenciones como unidades de medida, números, moneda, etc. No encontramos en nuestro corpus ningún error de este tipo, probablemente porque las unidades de medida en ambas lenguas coinciden y ningún segmento original incluía fechas concretas. Ello se debe a la naturaleza misma de los textos del corpus.

Indicador	Categoría	Valor del Error										Total Categoría					
		Addition		Omission		Mistranslation		Untranslated		Questionable			Misplaced		Register		Not idiomatic
		Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
1	Errores independientes de la combinación lingüística																
1.a	Errores de naturaleza lingüística																
1.a.1	Léxico extranjero					12	(12,50)	13	(13,54)								
1.a.2	Nombres propios					32	(33,33)	21	(21,88)								
1.b	Errores de naturaleza extralingüística																
1.b.1	Convenciones numéricas																
1.b.2	Elementos tipográficos	7	(7,29)	2	(2,08)							7	(7,29)				
1.b.3	Enlaces y vínculos	1	(1,04)			1	(1,04)										

Tabla 196: Número y porcentaje de errores marcados en la primera macrocategoría

En la ilustración a continuación podemos ver cómo se distribuyen los errores en las categorías de esta macrocategoría.

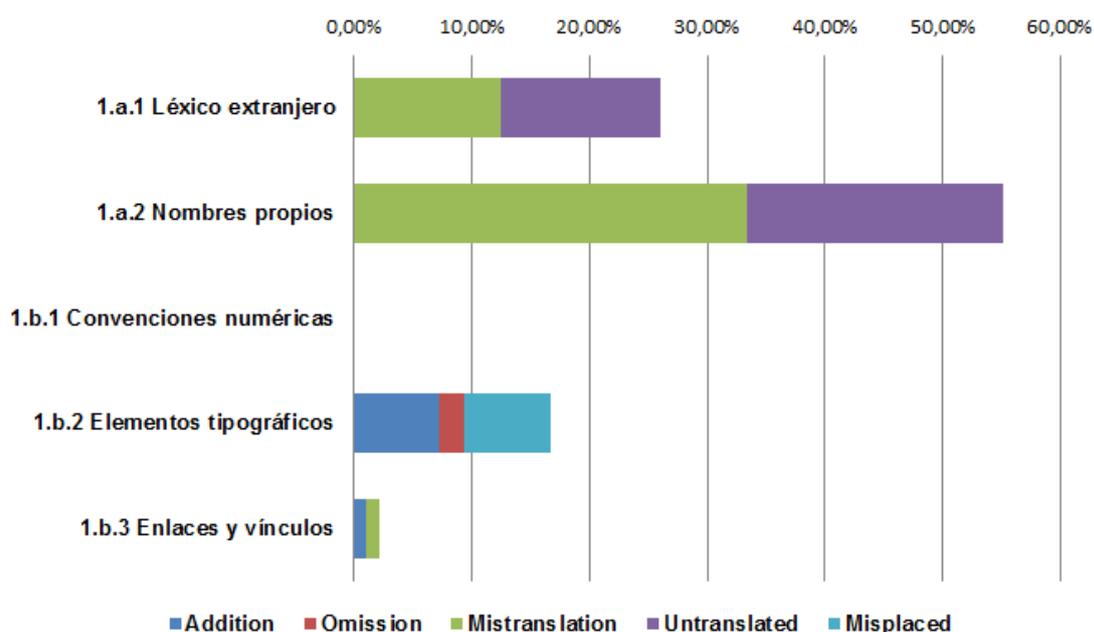


Ilustración 65: Porcentaje total de errores marcados en la primera macrocategoría

El hecho de que el porcentaje mayor de errores encontrados en las propuestas seleccionadas para la posesición tenga que ver con la traducción de nombres propios y con la presencia de léxico extranjero es un punto importante a tener en cuenta para la formulación de directrices de posesición. La mayoría de estas ediciones podrían evitarse o bien modificando el texto original o bien preparando los motores para este léxico específico.

9. Incidencia de cada categoría de errores en el corpus de análisis

9.1.2 Incidencia de los errores en la segunda macrocategoría

Siguiendo el mismo esquema del apartado anterior presentaremos en la tabla 181 (página 342) el porcentaje de errores que clasificamos como pertenecientes a la segunda macrocategoría, es decir aquellos errores provocados por las características propias del alemán. Aquí nos encontramos con que la mayoría de los errores se presentaron en la traducción de verbos compuestos separables; aquí marcamos 14 traducciones erróneas y en dos ocasiones se omitió el verbo. A continuación, encontramos errores en la traducción de conjunciones o partículas (10 errores), errores en la traducción de colocaciones o fraseologías (6 errores). Marcamos 3 errores en la reproducción de la negación. Encontramos 4 errores de transferencia sintáctica las composiciones, 2 por traducción errónea y 2 por omisión de la traducción.

Indicador	Categoría	Valor del Error										Total Categoría						
		Addition		Omission		Mistranslation		Untranslated		Questionable			Misplaced		Register		Not idiomatic	
		Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)		Núm.	(%)	Núm.	(%)	Núm.	(%)
2	Errores provocados por características propias del alemán																	
2.a	Verbos separables			2	(4,76)	14	(33,33)										16	
2.b	Conjunciones y partículas					10	(23,81)	1	(2,38)								11	
2.c	Composiciones			1	(2,38)	2	(4,76)										3	
2.d	Oraciones compuestas:																	
2.d.1	Oraciones subordinadas					2	(4,76)										2	
2.d.2	Oraciones coordinadas																	
2.e	Negación					3	(7,14)										3	
2.f	Colocaciones			1	(2,38)	6	(14,29)										7	

Tabla 197: Número y porcentaje de errores marcados en la segunda macrocategoría

La tabla 181 nos permite ver que la causa principal de errores en esta categoría se da en la traducción de verbos compuestos separables en alemán. Si sumamos los valores, es decir si sumamos los errores marcados sin tener en cuenta si fue un error de *mistranslation* o de *omission*, el porcentaje de errores es aún mayor en comparación con las otras categorías (38,06%). Ya en el marco teórico habíamos indicado que esta característica del alemán suele generar problemas en la traducción (ver página 69), especialmente en la traducción automática. El poseedor deberá tener en cuenta que esta característica propia del alemán es problemática y poner atención particularmente a la traducción de tales verbos. Del mismo modo debe tenerse en cuenta que los motores de traducción automática analizados también tienen problemas con la traducción de conjunciones y de colocaciones o fraseologías, ya que no siempre pueden encontrar el equivalente indicado. De nuevo, estos errores editados pueden también llevar a consideración el modo en que se redactan los textos originales para ser traducidos a través de traducción automática.

9. Incidencia de cada categoría de errores en el corpus de análisis

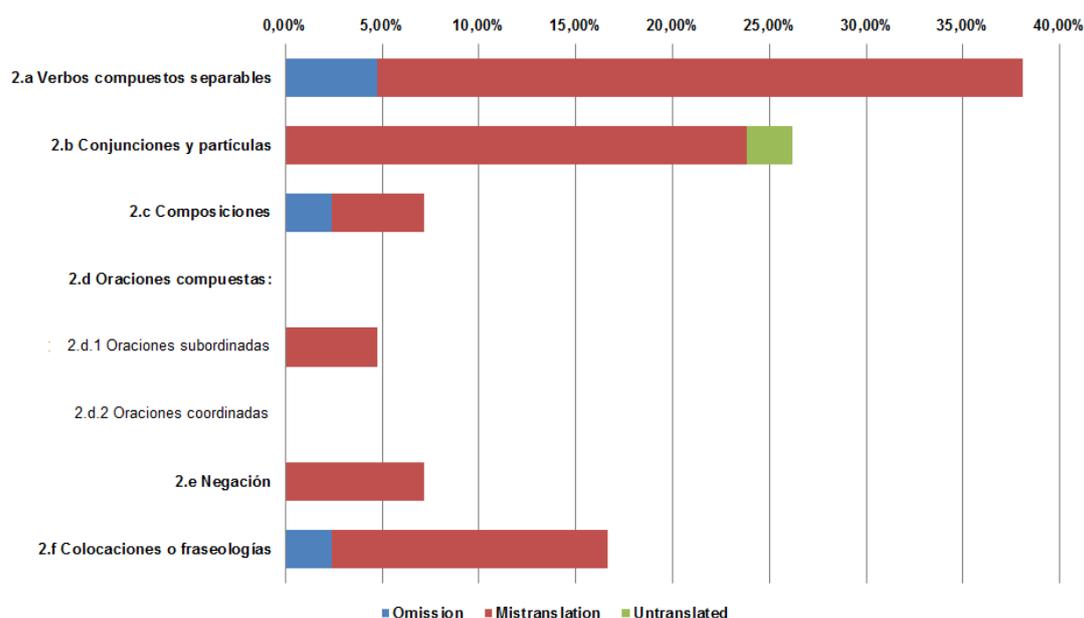


Ilustración 66: Porcentaje total de errores marcados en la segunda macrocategoría

9.1.3 Incidencia de los errores de la tercera macrocategoría

En la tercera macrocategoría en la que agrupamos las categorías de errores provocados por las características propias del castellano pudimos observar lo siguiente: el mayor porcentaje de errores lo constituyen los errores de orden sintáctico (182 errores). A continuación encontramos los errores en la flexión verbal (166 errores), seguidos de los errores ortográficos (95 errores). También se presentaron errores en la concordancia de número (42 errores) y de género (47 errores). Además, marcamos 29 errores tipográficos y 3 errores de estilo.

Indicador	Categoría	Valor del Error										Total Categoría						
		Addition		Omission		Mistranslation		Untranslated		Questionable			Misplaced		Register		Not idiomatic	
		Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)		Núm.	(%)	Núm.	(%)	Núm.	(%)
3	Errores provocados por características del castellano																	
3.a	Errores de flexión o concordancia																	
	3.a.1 Concordancia de género					47	(8,05)										47	
	3.a.2 Concordancia de número					62	(10,62)										62	
3.b	Errores de flexión verbal																	
	3.b.1 Flexión verbal (tiempo)					8	(1,37)										8	
	3.b.2 Flexión verbal (modo)					105	(17,98)										105	
	3.b.3 Flexión verbal (número)					41	(7,02)										41	
	3.b.4 Flexión verbal (voz)					12	(2,05)										12	
3.c	Estilo													1	(0,17)	2	(0,34)	
3.d	Errores ortotipográficos																	
	3.d.1 Ortografía	35	(5,99)	19	(3,25)					36	(6,16)	5	(0,86)				95	
	3.d.2 Tipografía	2	(0,34)	22	(3,77)					1	(0,17)	4	(0,68)				29	
3.e	Sintaxis	3	(0,51)	3	(0,51)							176	(30,14)				182	

Tabla 198: Número y porcentaje total de errores marcados en la tercera macrocategoría

La marcación de los errores en esta macrocategoría contiene una gran cantidad de detalle y por ello también tenemos una ilustración en la que vemos el porcentaje de errores de cada macrocategoría. Por ejemplo, pudimos diferenciar si los errores de sintaxis se dieron por

9. Incidencia de cada categoría de errores en el corpus de análisis

un ordenamiento erróneo de los elementos (*misplaced* 176 errores), por adición innecesaria de elementos (3 errores), por omisión de algún componente importante (1 error). También distinguimos qué tipo de error se dio en la flexión verbal y así pudimos determinar que el error más común aquí era la identificación del modo (105 errores).

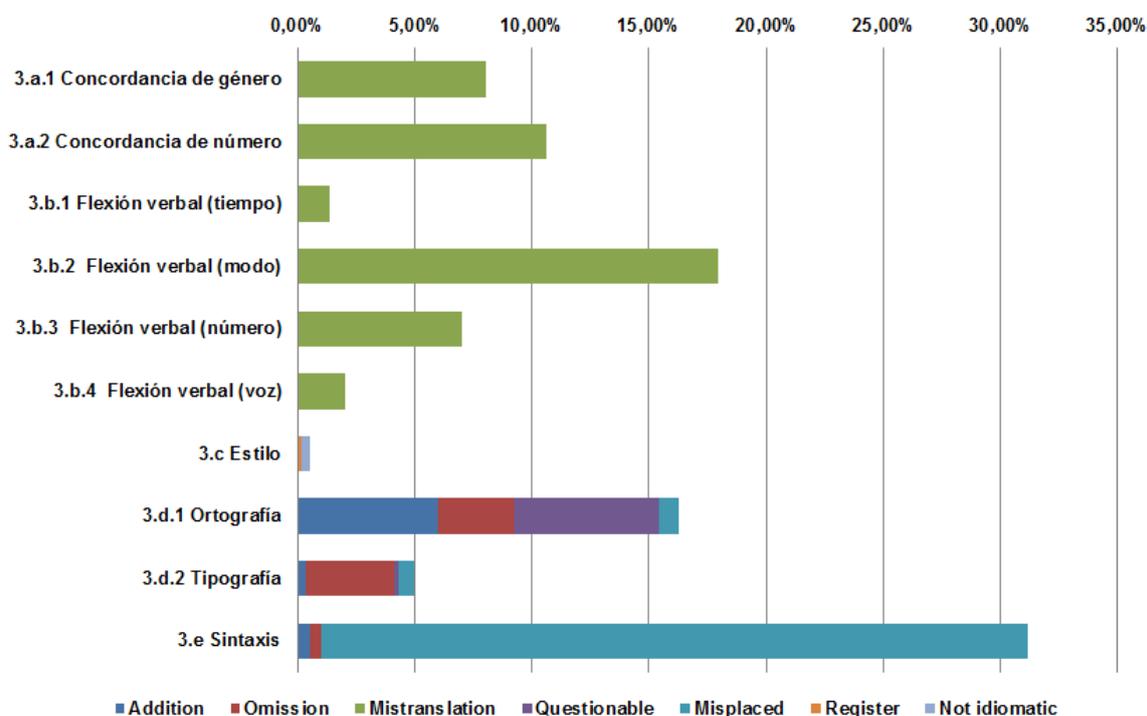


Ilustración 67: Porcentaje total de errores marcados en la tercera macrocategoría

Los datos nos indican que los motores de traducción automática reproducen generalmente la estructura sintáctica del alemán y por ello se presentan tantos errores de orden sintáctico, ya que la estructura del castellano es diferente. Los problemas de identificación del modo en castellano se dieron en particular porque la forma del imperativo y el infinitivo en alemán, así como la formulación de cortesía no se diferencian y por ello se tradujo generalmente como infinitivo y no se identificó el imperativo.

9.1.4 Incidencias de los errores de la cuarta macrocategoría

Como vimos anteriormente (página 312) en esta categoría marcamos el mayor número de errores. En la definición de las macrocategorías indicamos que la cuarta categoría se dividía en dos categorías principales: palabras con carga semántica y palabras sin carga semántica. Aquí presentaremos el porcentaje de encontrados teniendo también en cuenta esta división.

9. Incidencia de cada categoría de errores en el corpus de análisis

Indicador	Categoría	Valor del Error												Total Categoría				
		Addition		Omission		Mistranslation		Untranslated		Questionable		Misplaced			Register		Not idiomatic	
		Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)		Núm.	(%)	Núm.	(%)
4	Errores en la microestructura																	
4.a	Palabras sin carga semántica																	
4.a.1	Errores de carácter preposicional																	
4.a.1.1	Preposiciones	31	(2,92)	45	(4,23)	85	(8,00)	1	(0,09)									
4.a.1.2	Estructuras preposicionales	1	(0,09)	1	(0,09)	18	(1,69)											
4.a.2	Determinantes	68	(6,40)	68	(6,40)	12	(1,13)											
4.a.3	Errores de carácter adverbial																	
4.a.3.1	Adverbios	6	(0,56)	3	(0,28)	24	(2,26)											
4.a.3.2	Estructuras adverbiales					7	(0,66)											
4.a.4	Conjunciones y partículas	2	(0,19)	1	(0,09)													

Tabla 199: Porcentaje total de errores marcados en la cuarta macrocategoría en palabras sin carga semántica

En la ilustración anterior podemos ver que la mayoría de los errores se presentaron en la traducción de preposiciones, ya sea porque se añadieron (31 errores), se tradujeron erróneamente (85 errores), se omitieron (45 errores) o no se tradujeron (1 error). En segundo lugar encontramos un alto porcentaje de error en relación con los determinantes; en algunos casos se añadieron (68 errores), en otros se omitieron (también 68 casos) y en algunas ocasiones se seleccionó el equivalente adecuado (12 errores). La traducción de adverbios presenta menos errores: en 24 ocasiones no se tradujeron correctamente en 6 casos se añadieron y en 3 casos se omitieron.

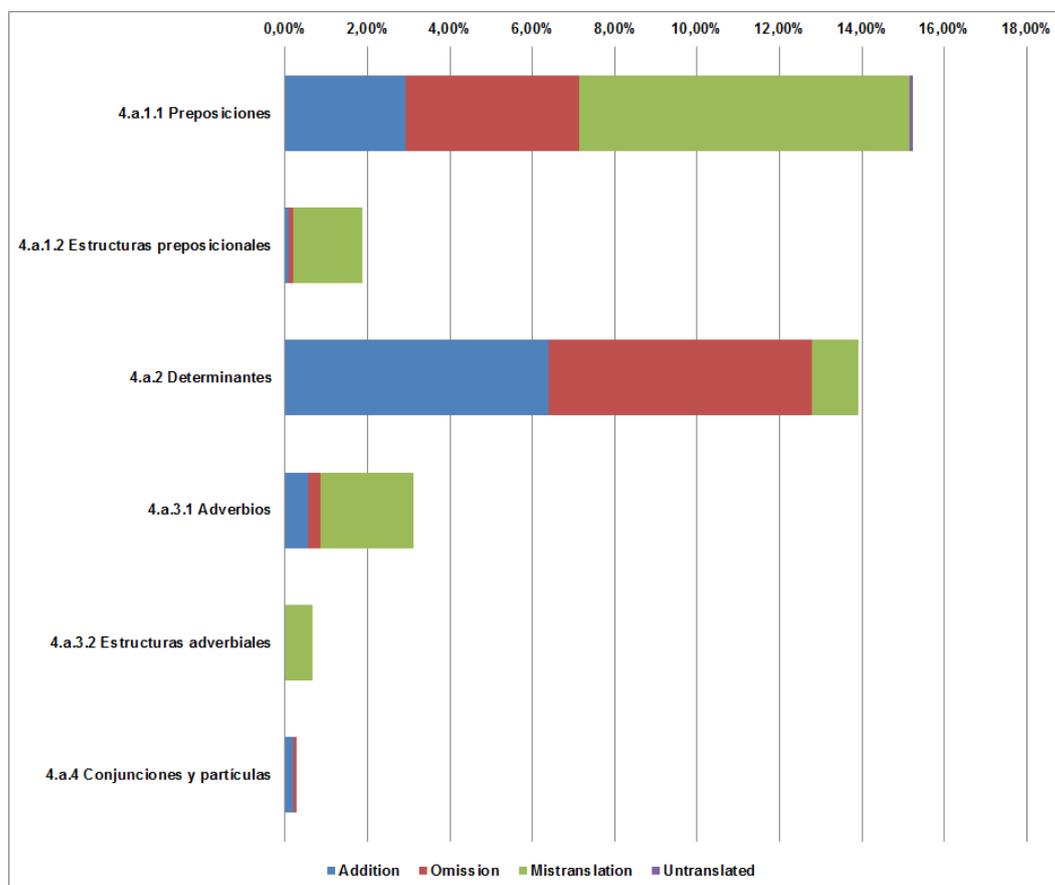


Ilustración 68: Porcentaje total de errores marcados en la cuarta macrocategoría en palabras sin carga semántica

9. Incidencia de cada categoría de errores en el corpus de análisis

En cuando a las palabras con carga semántica podemos observar que el mayor porcentaje de errores se da en la traducción de sustantivos y en la transferencia léxica de compuestos nominales y composiciones.

Indicador	Categoría	Valor del Error												Total Categoría				
		Addition		Omission		Mistranslation		Untranslated		Questionable		Misplaced			Register		Not idiomatic	
		Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)		Núm.	(%)	Núm.	(%)
4.b	Palabras con carga semántica																	
	4.b.1 Sustantivos	10	(0,94)	8	(0,75)	151	(14,21)	37	(3,48)								206	
	4.b.2 Adjetivos	4	(0,38)	8	(0,75)	53	(4,99)	8	(0,75)								73	
	4.b.3 Verbos																	
	4.b.3.1 Verbos simples	16	(1,51)	33	(3,10)	99	(9,31)	10	(0,94)			1	(0,09)				159	
	4.b.3.2 Verbos compuestos	1	(0,09)	10	(0,94)	41	(3,86)										52	
	4.b.4 Errores relativos a compuestos																	
	4.b.4.1 Compuestos nominales					42	(3,95)	4	(0,38)								46	
	4.b.4.2 Fluidez del sintagma					11	(1,03)										11	
	4.b.5 Transferencia léxica			3	(0,28)	116	(10,91)	24	(2,26)								143	

Tabla 200: Porcentaje total de errores marcados en la cuarta macrocategoría en palabras con carga semántica

Marcamos 151 ediciones por la traducción incorrecta de sustantivos, en 10 casos se añadieron en 8 ocasiones se omitieron y en 37 casos no se tradujeron. Lo mismo sucede con los adjetivos, la mayoría de los errores se marcaron porque no se encontró el equivalente correcto (53 errores), en 4 ocasiones se añadieron y en 8 ocasiones —respectivamente— se omitieron o se dejaron sin traducir. También se presentaron errores en la traducción de verbos simples (99 errores), en 33 casos se omitieron, en 10 no se tradujeron y en un caso se añadió un verbo innecesariamente. Menos errores se presentaron en la traducción de verbos compuestos: 41 errores de traducción, 1 adición y 10 omisiones. Un número mayor de errores se presentó en la transferencia léxica de compuestos nominales o composiciones 116 en total, en 3 ocasiones se omitieron y en 24 no se tradujeron.

9. Incidencia de cada categoría de errores en el corpus de análisis

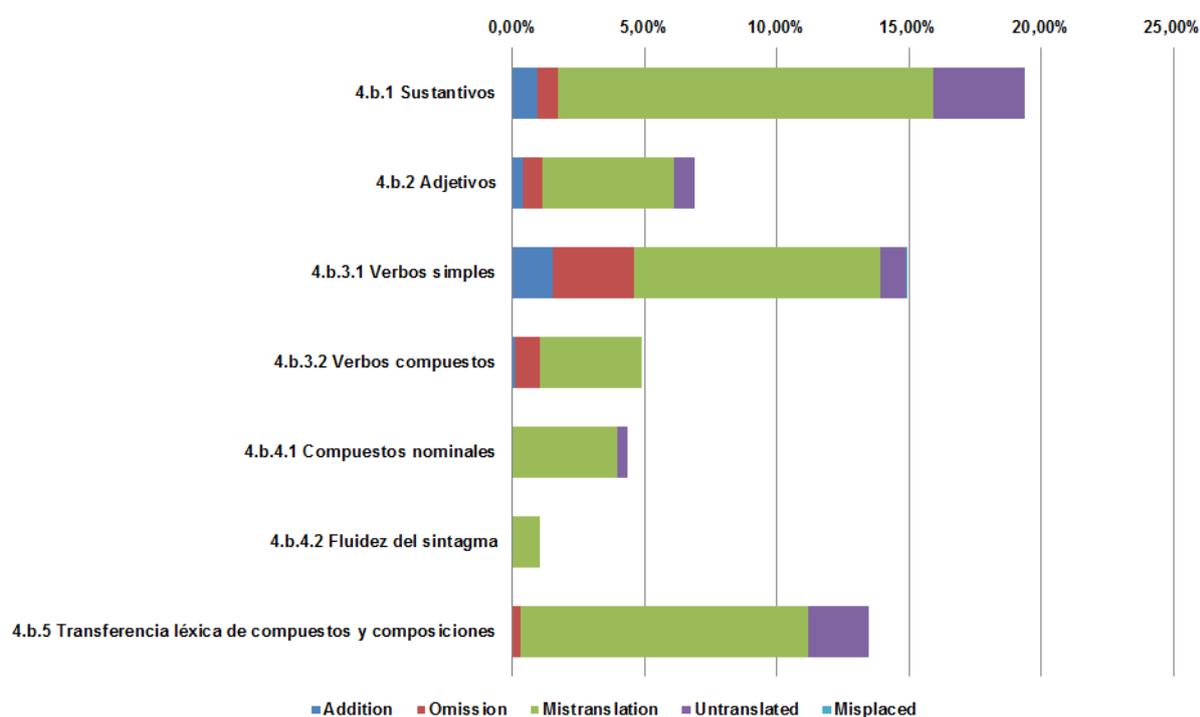


Ilustración 69: Porcentaje total de errores marcados en la cuarta macrocategoría en palabras con carga semántica

La traducción de los compuestos nominales (ver página 104) y composiciones en alemán al castellano es un aspecto problemático que hay que tener en cuenta para establecer reglas de posesición. Algunos de estos problemas se pueden evitar si el motor respectivo es entrenado con un diccionario especializado, pero los nombres compuestos y en particular las composiciones, como vimos en la página 307 son especialmente problemáticas porque no se encuentran en ningún diccionario y pueden ser creadas por el autor si el texto lo requiere. De todos modos, la categoría con mayor incidencia es la de los sustantivos, que no son objeto de traducción por su equivalente más apropiado. Así como la traducción de compuestos y composiciones es un problema tanto en la traducción con TA como sin ella, parece ser que la traducción de los sustantivos también lo pueda ser en el caso de la traducción con TA. En este caso probablemente se podría reducir la incidencia de este error mediante un control del léxico utilizado en el texto original, así como mediante una correcta gestión léxica dentro del motor.

9.1.5 Incidencia de los errores de la quinta macrocategoría

Como indicamos en el capítulo anterior, en esta macrocategoría incluimos todos los errores que tienen alguna relación con la redacción y diseño particulares del texto original.

9. Incidencia de cada categoría de errores en el corpus de análisis

Los errores más frecuentes en esta categoría se dan por la presencia de etiquetas en el texto original, que no son reconocidas como tales por el motor de traducción y por ello no se tratan correctamente. En la mayor parte de los casos se dañan (28 casos) y por ello las valoramos como *questionable*, en dos casos se colocan en la posición errada y en un caso se omiten. También se presentaron errores en la traducción por problemas en la redacción o estilo del segmento original (2 casos). Finalmente encontramos errores de tipo ortotipográfico que ocasionan que se traduzca erróneamente (8 errores); 2 casos en que se añaden y 5 casos en los que se omiten signos ortográficos relevantes para la comprensión del texto.

Indicador	Categoría	Valor del Error										Total Categoría						
		Addition		Omission		Mistranslation		Untranslated		Questionable			Misplaced		Register		Not idiomatic	
		Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)		Núm.	(%)	Núm.	(%)	Núm.	(%)
5	Errores relacionados con el texto original																	
5.a	Etiquetas			1	(2,08)					28	(58,33)	2	(4,17)					
5.b	Estilo															2	(4,17)	
5.c	Ortotipográfico	2	(4,17)	8	(16,67)	5	(10,42)											

Tabla 201: Número y porcentaje total de errores marcados en la quinta categoría

El análisis de los errores agrupados bajo esta categoría nos deja ver la importancia de que el texto original esté bien redactado y no presente errores. También es importante que las etiquetas sean marcadas o protegidas para evitar que se dañen. En caso de que esto no se cumpla el poseedor debe poder reconocer y solucionar este tipo de errores.

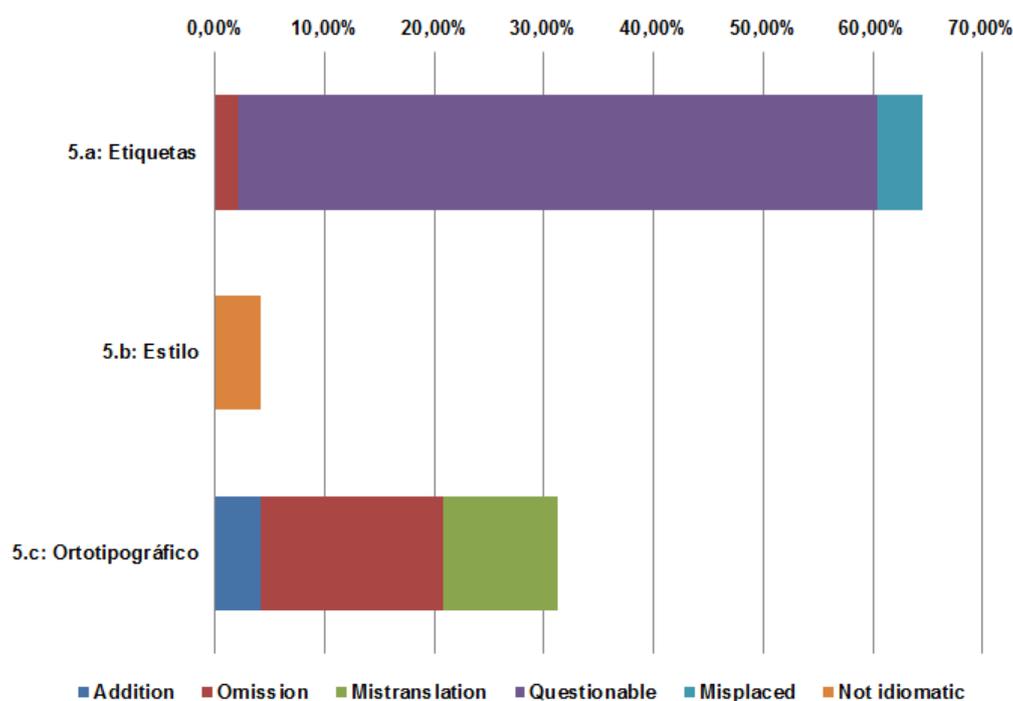


Ilustración 70: Porcentaje total de errores marcados en la quinta categoría

9. Incidencia de cada categoría de errores en el corpus de análisis

De nuevo, el problema de mayor incidencia de esta categoría podría minimizarse durante la fase de preparación del texto original o bien del sistema de traducción automática, si se invirtiera más esfuerzo en la revisión o preedición del texto original llevaría a un esfuerzo menor en la posesición, y esto es especialmente relevante cuando se traduce a varias lenguas.

9.1.6 Incidencia de los errores de la sexta macrocategoría

En la última categoría clasificamos los errores que no se podían ajustar a ninguna de las categorías mencionadas anteriormente. La totalidad de los 16 errores no tienen una explicación, es decir, no pudimos explicarnos porque se añadieron, por ejemplo, espacios en blanco innecesarios, palabras o combinaciones de palabras (5 errores) porque se tradujeron erróneamente ciertas oraciones (7 errores) o porque se omitieron palabras en la traducción (4 errores).

Indicador	Categoría	Valor del Error												Total Categoría				
		Addition		Omission		Mistranslation		Untranslated		Questionable		Misplaced			Register		Not idiomatic	
		Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)		Núm.	(%)	Núm.	(%)
6	Otro tipo de errores	5	(31,25)	4	(25,00)	7	(43,75)											16

Tabla 202: Número y porcentaje total de errores marcados en la sexta categoría

A continuación la ilustración de los resultados:

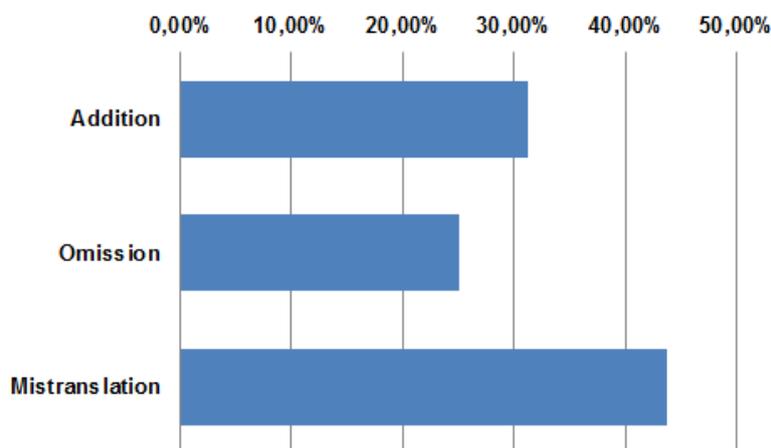


Ilustración 71: Porcentaje total de errores marcados en la sexta macrocategoría

Ya que este tipo de errores no tiene ninguna explicación aparente, tampoco pueden ayudarnos a formular estrategias para la posesición ni para la formación de poseedores.

9. Incidencia de cada categoría de errores en el corpus de análisis

Habiendo un recuento global de los errores encontrados en cada macrocategoría procederemos ahora a ver cómo es la incidencia de errores según el motor y si hay diferencias significativas dependiendo del tipo de motor.

9.2 Incidencia por macrocategoría con relación al tipo de motor

En este apartado no tendremos en cuenta los errores que marcamos en la categoría 6, es decir, los errores que no tienen justificación reconocible, ya que al no tener una explicación no nos ayudan a formular formas de sistematizar o de elaborar estrategias concretas para la posesición en nuestra combinación lingüística o para la formación de poseedores.

El número de segmentos editados con origen en cada motor es diferente. Por ello, indicar cuántos errores han sido editados en relación con cada motor podría inducir a una percepción errónea, dado que el motor cuyos segmentos prefirieron los poseedores presentó un mayor número de ediciones por haber sido el que ha aportado más segmentos al corpus. Por este motivo los datos se presentarán de manera ponderada. En concreto realizaremos una doble ponderación: en primer lugar, en función del número de segmentos editados de cada sistema o motor de traducción automática, y en segundo lugar en función de la cantidad de errores editados en la macrocategoría o categoría analizada en ese momento. Esta doble ponderación no se debe entender como una escala cerrada sino que se trata únicamente de un valor que nos permitirá evaluar el comportamiento de los motores en términos relativos entre sí.

En la siguiente ilustración podemos ver el porcentaje total de errores marcados por macrocategoría teniendo en cuenta el tipo de sistema. Aquí tenemos que volver a mencionar, que la mayoría de las propuestas seleccionadas para la posesición fueron las propuestas del RBMT1 y por ello aparecen marcados muchos más errores (ver ilustración en la página 311). Ello no quiere decir, como también vemos en la ilustración, que el tipo de sistema que comete más errores sea el sistema basado en reglas, sino que por tener más propuestas pudimos analizar con más detalle los errores que se presentan en este tipo de motores. Además, también cabe señalar que el sistema con más segmentos no debe interpretarse como el motor que cometió más errores, sino como el motor cuyos segmentos prefirieron utilizar los poseedores por delante de los segmentos de los otros tres sistemas. Por lo que, a priori, se trata de los segmentos más simples de poseer de los cuatro posibles.

9. Incidencia de cada categoría de errores en el corpus de análisis

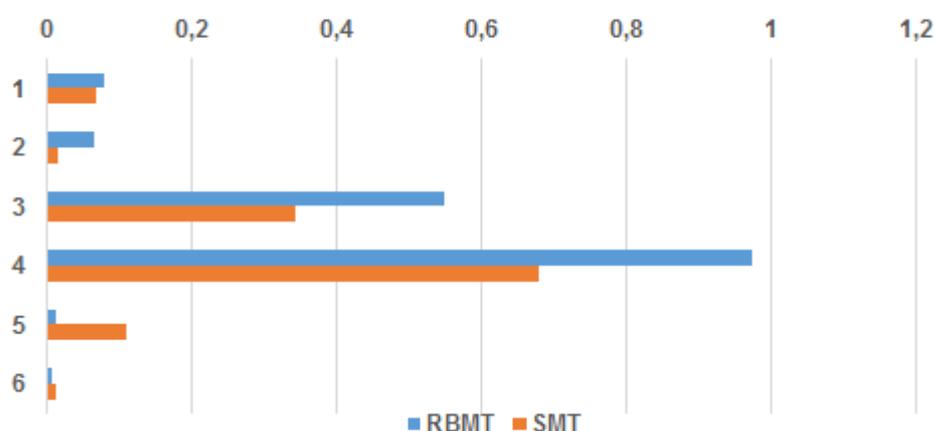


Ilustración 72: Número de errores marcados clasificados por macrocategorías según el motor seleccionado

En la tabla a continuación anotamos el número de errores marcados por tipo de motor. Como vimos en el análisis preliminar del corpus, en las propuestas de ciertos motores se presenta más algunos tipos de errores que otros (ver capítulo “Errores encontrados según el motor empleado” a partir de la página 255). En este capítulo analizaremos como es la relación o incidencia de las macrocategorías y categorías definidas en el capítulo anterior a cada tipo de motor. Para ello primero tenemos que tener en cuenta el número de errores marcados por tipo de sistema y por categoría.

Tipo de motor	Primera macrocategoría	Segunda macrocategoría	Tercera macrocategoría	Cuarta macrocategoría	Quinta macrocategoría
SMT	35	13	193	369	32
RBMT	59	40	382	691	17

Tabla 203: Número total de segmentos seleccionados por macrocategoría por tipo de sistema

A primera vista y sin entrar todavía en detalle podemos ver que la categoría en la que se marcaron más errores, tanto en los motores estadísticos como en los motores basados en reglas, fue la cuarta categoría. Esto parece indicar que los poseedores prefieren editar problemas de microestructura (cuarta macrocategoría) o prefirieron editar cuestiones errores de flexión, concordancia o conjugación de verbos (tercera macrocategoría) que otros posibles aspectos. En los apartados a continuación analizaremos detalladamente la relación de cada tipo de motor y cada categoría.

9. Incidencia de cada categoría de errores en el corpus de análisis

9.2.1 Incidencia por tipos de motor en la primera macrocategoría

Empezaremos por ver cómo es la relación de errores marcados en las categorías con respecto a los motores estadísticos en cada macrocategoría y en sus subcategorías.

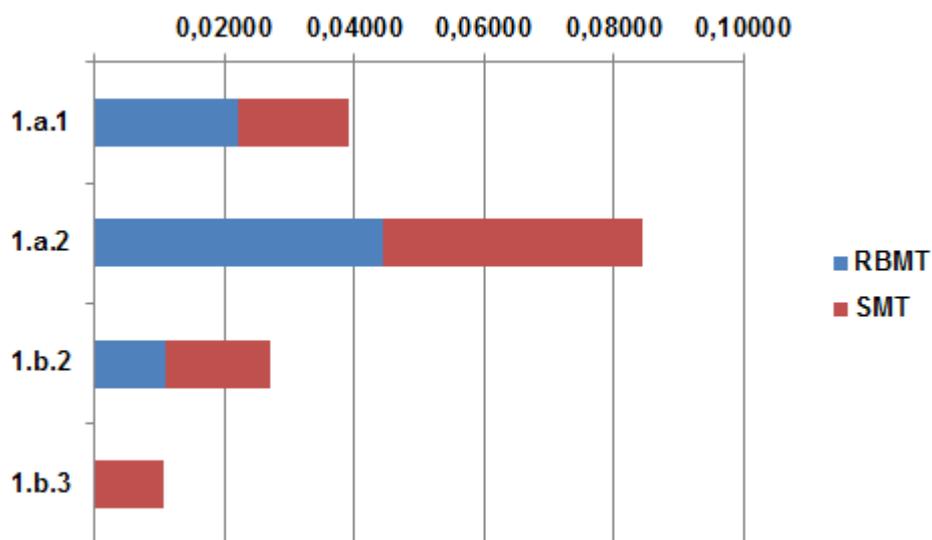


Ilustración 73: Incidencia de los tipos de motores en las categorías de la primera macrocategoría

En la tabla a continuación se encuentran listados los errores por categoría y por valor asignado. De acuerdo a estos datos los poseedores editaron en esta categoría errores en la traducción de nombres propios y en la traducción o adaptación de léxico extranjero. También es interesante el número de ediciones que se tuvieron que realizarse ediciones en los casos en que los nombres propios o palabras en léxico extranjero no fueron traducidos, cuando se debían traducir.

smt	Categoría	Valor del Error															
		Addition		Omission		Mistranslation		Untranslated		Questionable		Misplaced		Register		Not idiomatic	
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
1	Errores independientes de la combinación lingüística																
1.a	Errores de naturaleza lingüística																
1.a.1	Léxico extranjero					7	(7,29)	2	(2,08)								
1.a.2	Nombres propios					11	(11,46)	9	(9,38)								
1.b	Errores de naturaleza extralingüística																
1.b.1	Convenciones numéricas																
1.b.2	Elementos tipográficos	6	(6,25)									1	(1,04)				
1.b.3	Enlaces y vínculos	1	(1,04)			1	(1,04)										

Ilustración 74: Errores encontrados en la primera macrocategoría SMT

Aunque la diferencia no es muy grande, según los porcentajes que vemos en la tabla siguiente, los editores que se decidieron por los segmentos de los motores basados en reglas

9. Incidencia de cada categoría de errores en el corpus de análisis

prefieren editar una palabra que no se tradujo correctamente (26 casos) a traducir una palabra que se dejó de traducir (23 casos).

rbmt	Categoría	Valor del Error															
		Addition		Omission		Mistranslation		Untranslated		Questionable		Misplaced		Register		Not idiomatic	
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
1	Errores independientes de la combinación lingüística																
1.a	Errores de naturaleza lingüística																
1.a.1	Léxico extranjero					5	(5,21)		11	(11,46)							
1.a.2	Nombres propios					21	(21,88)		12	(12,50)							
1.b	Errores de naturaleza extralingüística																
1.b.1	Convenciones numéricas																
1.b.2	Elementos tipográficos	1	(1,04)	2	(2,08)							6	(6,25)				
1.b.3	Enlaces y vínculos																

Ilustración 75: Errores encontrados en la primera macrocategoría RBMT

Las otras categorías de errores o bien no se dieron a partir de los originales del corpus o bien pudieron aparecer en traducciones que los poseedores no utilizaron.

9.2.2 Incidencia por tipos de motor en la segunda macrocategoría

En la ilustración a continuación comparamos los resultados de las ediciones que se hicieron cuando se presentaron errores de la segunda categoría, es decir, aquellos provocados por las características del alemán según los tipos de motores:

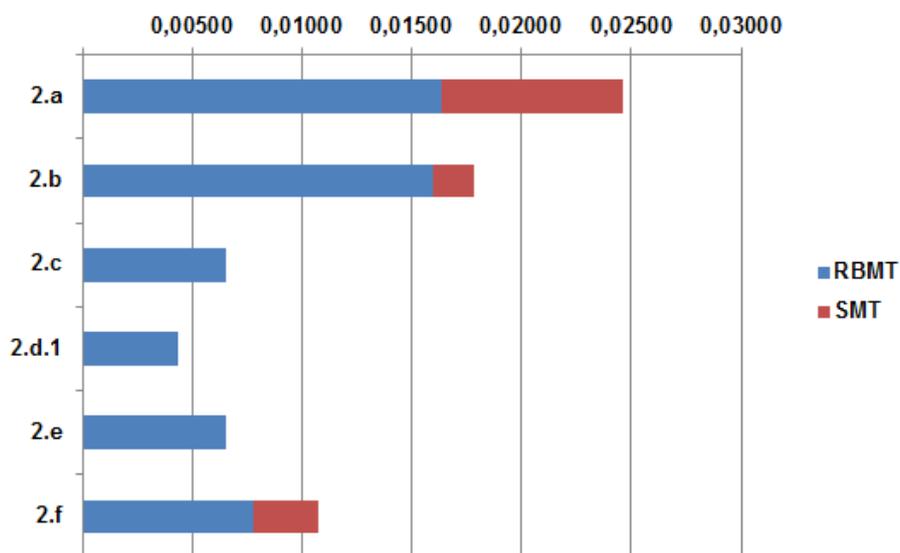


Ilustración 76: Incidencia de los tipos de motores en las categorías de la segunda macrocategoría

Como podemos ver la mayor cantidad de ediciones se realizó cuando se presentaron errores en la traducción de verbos separables. Este es un problema a tener en cuenta para la

9. Incidencia de cada categoría de errores en el corpus de análisis

posedición en nuestra combinación lingüística, ya que ambos tipos de motores presentan problemas en la traducción de estas palabras.

smt	Categoría	Valor del Error															
		Addition		Omission		Mistranslation		Untranslated		Questionable		Misplaced		Register		Not idiomatic	
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
2	Errores provocados por características propias del ale																
2.a	Verbos separables			2	(4,76)	3	(7,14)										
2.b	Conjunciones y partículas					2	(4,76)										
2.c	Composiciones																
2.d	Oraciones compuestas:																
	2.d.1 Oraciones subordinadas																
	2.d.2 Oraciones coordinadas																
2.e	Negación																
2.f	Colocaciones			1	(2,38)	1	(2,38)										

Ilustración 77: Errores encontrados en la segunda macrocategoría SMT

En la tabla a continuación podemos ver que en las propuestas más seleccionadas se refleja proporcionalmente lo indicado anteriormente. Los poseedores prefirieron editar problemas de transferencia de la negación (5 casos) y problemas en la traducción de oraciones compuestas. Al igual que en los motores estadísticos el mayor número de ediciones se realizó en la traducción de verbos separables.

rbmt	Categoría	Valor del Error															
		Addition		Omission		Mistranslation		Untranslated		Questionable		Misplaced		Register		Not idiomatic	
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
2	Errores provocados por características propias del ale																
2.a	Verbos separables					11	(26,19)										
2.b	Conjunciones y partículas					8	(19,05)		1	(2,38)							
2.c	Composiciones			1	(2,38)	2	(4,76)										
2.d	Oraciones compuestas:																
	2.d.1 Oraciones subordinadas					2	(4,76)										
	2.d.2 Oraciones coordinadas																
2.e	Negación					3	(7,14)										
2.f	Colocaciones					5	(11,90)										

Ilustración 78: Errores encontrados en la segunda macrocategoría RBMT

Ahora veremos qué relación se da entre los tipos de motores y los problemas ocasionados por características propias del castellano.

9. Incidencia de cada categoría de errores en el corpus de análisis

9.2.3 Incidencia por tipos de motor en la tercera macrocategoría

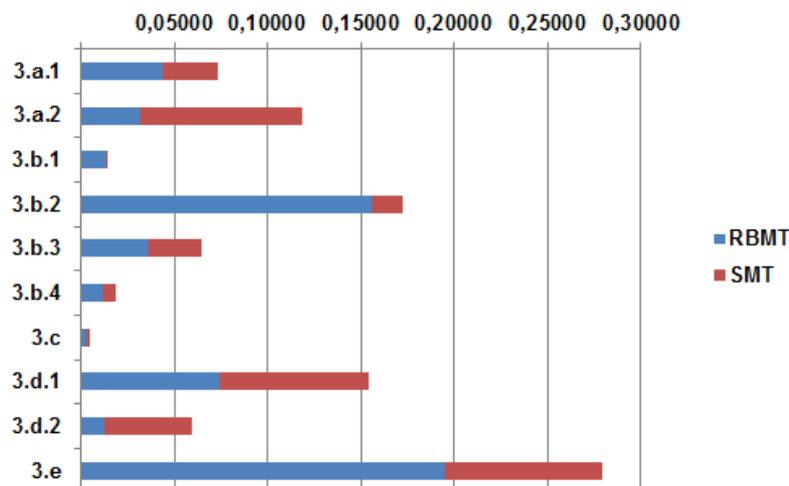


Ilustración 79: Incidencia de los tipos de motores en las categorías de la tercera macrocategoría

En la ilustración anterior podemos ver que se editaron más errores de orden sintáctico en las traducciones de los sistemas basados en reglas que en las de los sistemas estadísticos.

En esta categoría en las propuestas de los motores estadísticos seleccionados el mayor número de ediciones se hicieron por errores de sintaxis (54 casos), por errores de ortografía (38 casos) y por ajustes en la concordancia de número (32 casos marcados).

smt	Valor del Error	Valor del Error							
		Addition	Omission	Mistranslation	Untranslated	Questionable	Misplaced	Register	Not idiomatic
3	Errores provocados por características del castellano								
3.a	Errores de flexión o concordancia								
3.a.1	Concordancia de género			16 (2,74)					
3.a.2	Concordancia de número			32 (5,48)					
3.b	Errores de flexión verbal								
3.b.1	Flexión verbal (tiempo)			1 (0,17)					
3.b.2	Flexión verbal (modo)			18 (3,08)					
3.b.3	Flexión verbal (número)			15 (2,57)					
3.b.4	Flexión verbal (voz)			4 (0,68)					
3.c	Estilo								1 (0,17)
3.d	Errores ortotipográficos								
3.d.1	Ortografía	19 (3,25)	11 (1,88)			6 (1,03)	2 (0,34)		
3.d.2	Tipografía	1 (0,17)	12 (2,05)			1 (0,17)	2 (0,34)		
3.e	Sintaxis	3 (0,51)	2 (0,34)				49 (8,39)		

Ilustración 80: Errores encontrados en la tercera macrocategoría SMT

En esta categoría en las propuestas de los motores basados en reglas seleccionados el mayor número de ediciones también se hicieron por errores de sintaxis (128 casos), por errores de ortografía (57 casos) y por problemas en la conjugación verbal 128 casos marcados).

9. Incidencia de cada categoría de errores en el corpus de análisis

rbmt		Valor del Error							
		Addition	Omission	Mistranslation	Untranslated	Questionable	Misplaced	Register	Not idiomatic
3	Errores provocados por características del castellano								
3.a	Errores de flexión o concordancia								
3.a.1	Concordancia de género			31 (5,31)					
3.a.2	Concordancia de número			30 (5,14)					
3.b	Errores de flexión verbal								
3.b.1	Flexión verbal (tiempo)			7 (1,20)					
3.b.2	Flexión verbal (modo)			87 (14,90)					
3.b.3	Flexión verbal (número)			26 (4,45)					
3.b.4	Flexión verbal (voz)			8 (1,37)					
3.c	Estilo							1 (0,17)	1 (0,17)
3.d	Errores ortotipográficos								
3.d.1	Ortografía	16 (2,74)	8 (1,37)			30 (5,14)	3 (0,51)		
3.d.2	Tipografía	1 (0,17)	10 (1,71)				2 (0,34)		
3.e	Sintaxis		1 (0,17)						

Ilustración 81: Errores encontrados en la tercera macrocategoría RBMT

Si bien los problemas se dan en una proporción diferente se puede ver cierta tendencia relativa a que la mayor cantidad de ediciones se hizo en la corrección del orden sintáctico, ortográfico o de la flexión de número o género.

9.2.4 Incidencia por tipos de motor en la cuarta macrocategoría

Debido a que esta macrocategoría es tan amplia analizaremos los resultados separando las dos categorías principales, es decir, nos referiremos primero a los errores que tienen que ver con las palabras sin carga semántica. La mayor cantidad de ediciones se realizaron en la corrección de la traducción de preposiciones o estructuras preposicionales. En segundo lugar, se editaron las traducciones de determinantes y en tercer lugar los adverbios.

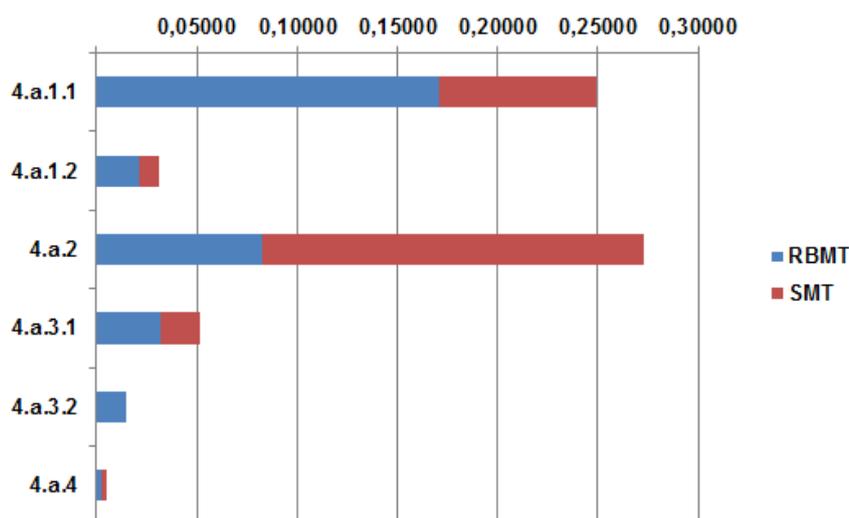


Ilustración 82: Incidencia de los tipos de motores en las categorías de la cuarta macrocategoría sin carga semántica

Encontramos aquí la segunda categoría en la que se prefirió poseer segmentos procedentes de SMT en lugar de RBMT, la categoría 4.a.2.

9. Incidencia de cada categoría de errores en el corpus de análisis

Ahora veamos lo que ocurre con las palabras con carga semántica. Aquí el mayor número de ediciones se encontró en la traducción de sustantivos, seguidos por la traducción de verbos y la transferencia de compuestos nominales.

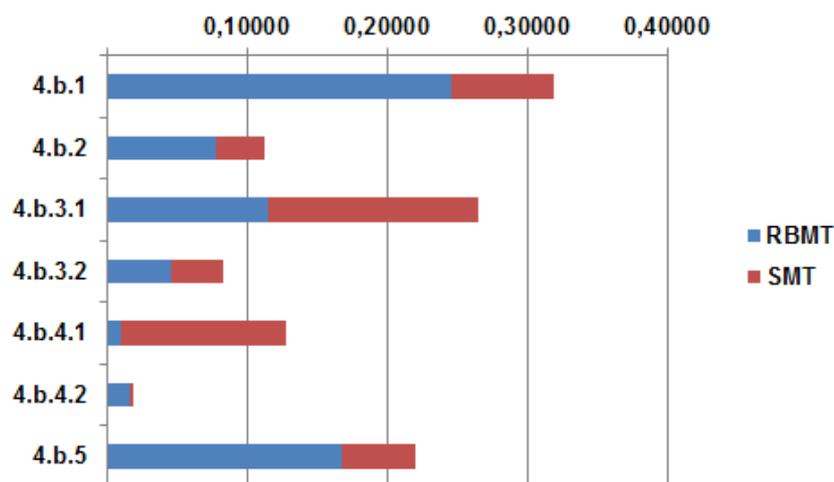


Ilustración 83: Incidencia de los tipos de motores en las categorías de la cuarta macrocategoría con carga semántica

En este bloque hay dos categorías en las que se prefirió poseer segmentos procedentes de SMT: la 4.b.3.1. y la 4.b.4.1.

Si vemos separadamente los resultados de cada tipo de motor, en los motores estadísticos se editaron las adiciones de determinantes (45 casos), las traducciones incorrectas de sustantivos (38 casos) y de preposiciones (26 casos).

smt	Error en la microestructura	Valor del Error							
		Addition	Omission	Mistranslation	Untranslated	Questionable	Misplaced	Register	Not idiomatic
4	Errores en la microestructura								
4.a	Palabras sin carga semántica								
4.a.1	Errores de carácter preposicional								
4.a.1.1	Preposiciones	9 (0,85)	14 (1,32)	26 (2,45)					
4.a.1.2	Estructuras preposicionales			6 (0,56)					
4.a.2	Determinantes	45 (4,23)	21 (1,98)	7 (0,66)					
4.a.3	Errores de carácter adverbial								
4.a.3.1	Adverbios	3 (0,28)	3 (0,28)	5 (0,47)					
4.a.3.2	Estructuras adverbiales								
4.a.4	Conjunciones y partículas	1 (0,09)							
4.b	Palabras con carga semántica								
4.b.1	Sustantivos	6 (0,56)	3 (0,28)	38 (3,57)	6 (0,56)				
4.b.2	Adjetivos	4 (0,38)	6 (0,56)	8 (0,75)	4 (0,38)				
4.b.3	Verbos								
4.b.3.1	Verbos simples	10 (0,94)	31 (2,92)	24 (2,26)	2 (0,19)				
4.b.3.2	Verbos compuestos		9 (0,85)	10 (0,94)					
4.b.4	Errores relativos a compuestos								
4.b.4.1	Compuestos nominales			28 (2,63)	4 (0,38)				
4.b.4.2	Fluidez del sintagma			2 (0,19)					
4.b.5	Transferencia léxica		2 (0,19)	23 (2,16)	13 (1,22)				

Ilustración 84: Errores encontrados en la cuarta macrocategoría SMT

En los motores basados en reglas la proporción es varía levemente. Se editaron en 59 casos traducciones incorrectas de preposiciones. Pero el mayor número de ediciones se da en

9. Incidencia de cada categoría de errores en el corpus de análisis

la corrección de la traducción de sustantivos (113 casos) y en la edición de errores de transferencia léxica de compuestos nominales (93 casos).

rbmt		Valor del Error							
		Addition	Omission	Mistranslation	Untranslated	Questionable	Misplaced	Register	Not idiomatic
4	Errores en la microestructura								
4.a	Palabras sin carga semántica								
4.a.1	Errores de carácter preposicional								
4.a.1.1	Preposiciones	22 (2,07)	31 (2,92)	59 (5,55)	1 (0,09)				
4.a.1.2	Estructuras preposicionales	1 (0,09)	1 (0,09)	12 (1,13)					
4.a.2	Determinantes	23 (2,16)	47 (4,42)	5 (0,47)					
4.a.3	Errores de carácter adverbial								
4.a.3.1	Adverbios	3 (0,28)		19 (1,79)					
4.a.3.2	Estructuras adverbiales			7 (0,66)					
4.a.4	Conjunciones y partículas	1 (0,09)	1 (0,09)						
4.b	Palabras con carga semántica								
4.b.1	Sustantivos	4 (0,38)	5 (0,47)	113 (10,63)	31 (2,92)				
4.b.2	Adjetivos		2 (0,19)	45 (4,23)	4 (0,38)				
4.b.3	Verbos								
4.b.3.1	Verbos simples	6 (0,56)	2 (0,19)	75 (7,06)	8 (0,75)		1 (0,09)		
4.b.3.2	Verbos compuestos	1 (0,09)	1 (0,09)	31 (2,92)					
4.b.4	Errores relativos a compuestos								
4.b.4.1	Compuestos nominales			14 (1,32)					
4.b.4.2	Fluidez del sintagma			9 (0,85)					
4.b.5	Transferencia léxica		1 (0,09)	93 (8,75)	11 (1,03)				

Ilustración 85: Errores encontrados en la cuarta macrocategoría RBMT

Los poseedores prefieren corregir problemas de transferencia léxica o de traducción incorrecta de sustantivos. Del mismo modo parece una tendencia que es más fácil traducir palabras que quedaron sin traducir de cero, que corregirlas (55 casos en los motores basados en reglas y 29 en los motores estadísticos).

9.2.5 Incidencia por tipos de motor en la quinta macrocategoría

Como lo indicamos anteriormente en esta categoría anotamos los errores que tenían que ver con el original, es decir, aquellos que se dan por problemas en los segmentos a traducir. El problema principal lo constituye la presencia de etiquetas.

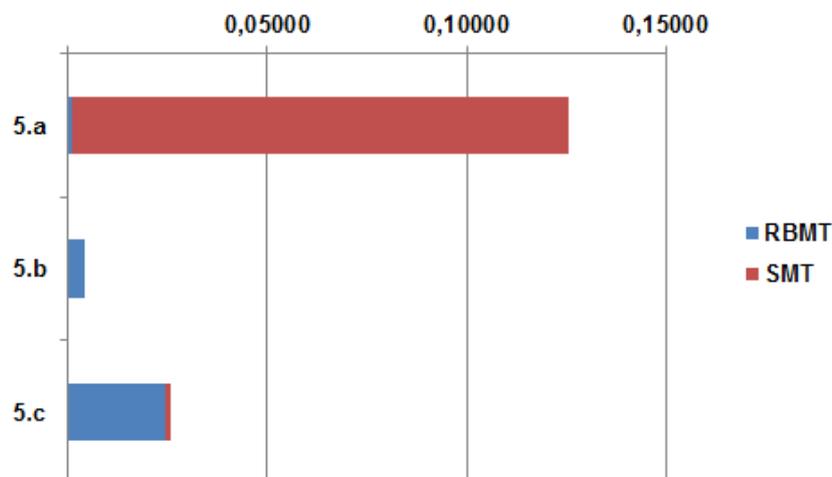


Ilustración 86: Incidencia de los tipos de motores en las categorías de la quinta macrocategoría

9. Incidencia de cada categoría de errores en el corpus de análisis

La categoría 5.a tuvo incidencia mayoritaria en segmentos procedentes de SMT. Los motores estadísticos tuvieron más problemas con el tratamiento de las etiquetas. Esta es una de las causas porque sus propuestas no se escogieron con tanta frecuencia. También debemos recordar que en el análisis preliminar del corpus detectamos que había segmentos con errores con etiquetas entre los no utilizados para poseer que, de no presentar el error de la etiqueta habrían sido preferidos por el poseedor (ver apartados correspondientes a cada motor estadístico en la página 255 y en la página 257).

smt		Valor del Error							
		Addition	Omission	Mistranslation	Untranslated	Questionable	Misplaced	Register	Not idiomatic
5	Errores relacionados con el texto original								
5.a	Etiquetas					27 (56,25)			
5.b	Estilo								
5.c	Ortotipográfico			2 (4,17)					
6	Otro tipo de errores	5 (31,25)	3 (18,75)	1 (6,25)					

Ilustración 87: Errores encontrados en la quinta macrocategoría SMT

En las propuestas de los motores basados en reglas no se realizaron tantas ediciones en esta categoría. Y aquellas que se realizaron se dieron por problemas de orden ortotipográfico en el original. Esto probablemente se deba a que el error en el tratamiento de etiquetas no se dé en segmentos procedentes de RBMT.

rbmt		Valor del Error							
		Addition	Omission	Mistranslation	Untranslated	Questionable	Misplaced	Register	Not idiomatic
5	Errores relacionados con el texto original								
5.a	Etiquetas		1 (2,08)			1 (2,08)	2 (4,17)		
5.b	Estilo								2 (4,17)
5.c	Ortotipográfico	2 (4,17)	8 (16,67)	3 (6,25)					
6	Otro tipo de errores		1 (6,25)	6 (37,50)					

Ilustración 88: Errores encontrados en la quinta macrocategoría RBMT

A pesar de que hay similitud entre los motores queremos ver si cada motor por separado se comporta de una manera diferente.

9.3 Incidencia por categorías y motor concreto

Habiendo analizado los errores encontrados en su totalidad y según el tipo de error, ahora veremos cómo es la relación en cada motor concreto. Ya en la primera fase del análisis vimos que hay cierto comportamiento similar de los motores en relación con algunos tipos de errores; sin embargo, también pudimos darnos cuenta que hay diferencias entre los motores mismos, a pesar de que sean del mismo tipo. En este apartado veremos cuáles son estas diferencias. Por ello partiremos de cada macrocategoría y compararemos los errores encontrados

9. Incidencia de cada categoría de errores en el corpus de análisis

en cada tipo de motor. Como vimos en la ilustración en la página 312 el número de las propuestas seleccionadas para la posesición según el motor no es igual: de todos los 651 segmentos analizados 111 eran propuestas realizadas por el motor SMT1, 78 del motor SMT2, 411 del RBMT1 y 51 propuestas del motor RBMT2. Por ello hemos realizado un cálculo con doble ponderación para realizar una comparación más exacta de los datos encontrados en el corpus. En la siguiente ilustración vemos los resultados:

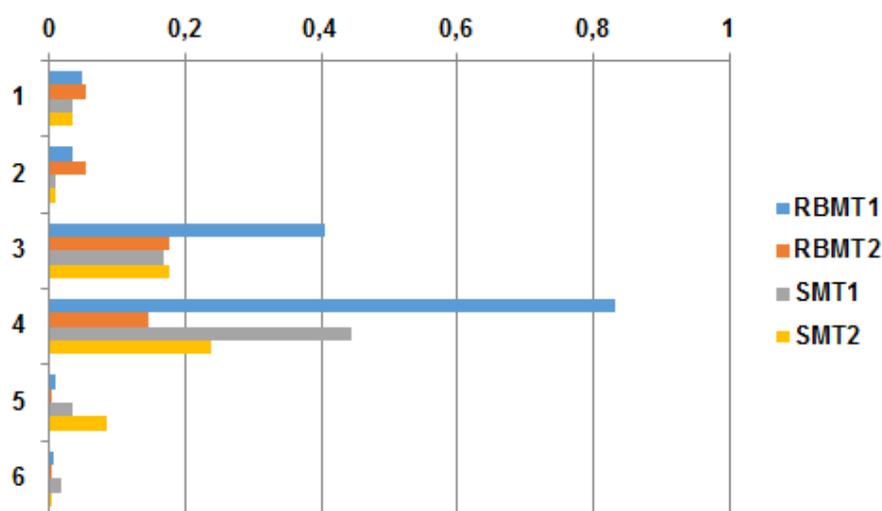


Ilustración 89: Relación ponderada de los errores encontrados en cada macrocategoría y en cada motor

Aquí podemos ver que en los segmentos procedentes de los motores RBMT1, SMT1 y SMT2 los errores más frecuentes se dan a nivel de la microestructura; solo el motor RBMT2 tuvo más problemas en solucionar los problemas propios del castellano. Como en el apartado anterior analizaremos ahora detalladamente solo las cinco primeras macrocategorías.

9.3.1 Primera macrocategoría: errores independientes de la combinación lingüística

En la ilustración a continuación podemos observar que los motores basados en reglas tuvieron menos problemas en interpretar la funcionalidad de elementos tipográficos (categoría 1.b.2) y con las referencias, como enlaces, vínculos e hipertexto (categoría 1.b.3). Pero en lo que tiene que ver con la presencia de léxico extranjero observamos dos fenómenos interesantes: si bien los motores estadísticos cometen más errores, los motores basados en reglas dejan en un porcentaje más alto los extranjerismos sin traducir.

9. Incidencia de cada categoría de errores en el corpus de análisis

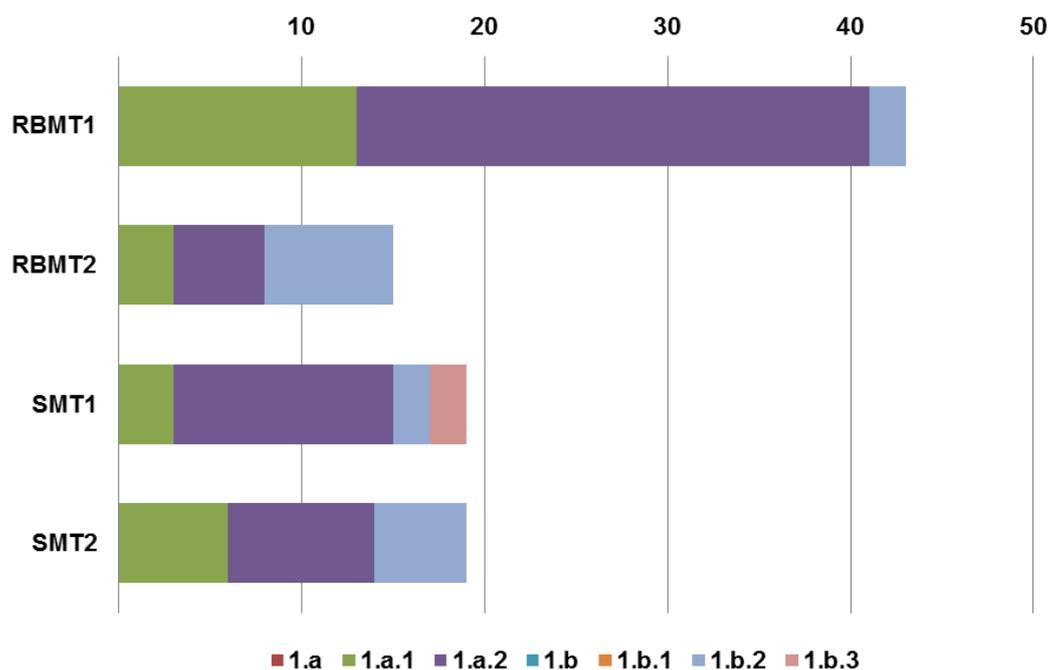


Ilustración 90: Relación por motor en las categorías de la primera macrocategoría

Para ver la incidencia de los motores los veremos en primer lugar en cada categoría y por separado, para observar si hay diferencias o similitudes entre ellos. Empezaremos con la primera categoría. En la tabla siguiente, podemos observar el número de errores marcados en cada categoría y el porcentaje de estos con relación al total de ediciones realizadas en la macrocategoría (96 ediciones).

smt1	Indicador	Categoría	Valor del Error								Total Categoría
			Addition		Omission		Mistranslation		Untranslated		
			Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
1		Errores independientes de la combinación lingüística									
	1.a	Errores de naturaleza lingüística									
	1.a.1	Léxico extranjero					2	(2,08)	1	(1,04)	3
	1.a.2	Nombres propios					9	(9,38)	3	(3,13)	12
	1.b	Errores de naturaleza extralingüística									
	1.b.1	Convenciones numéricas									
	1.b.2	Elementos tipográficos	2	(2,08)							2
	1.b.3	Enlaces y vínculos	1	(1,04)			1	(1,04)			2

Tabla 204: Número y porcentaje total de errores marcados en la primera macrocategoría en el motor SMT1

Aquí podemos ver que el error más frecuente es la traducción incorrecta de nombres propios. En nueve casos se tradujeron erróneamente y en tres casos se dejaron de traducir. En cuando a las palabras en léxico extranjero se marcaron dos casos de *mistranslation* y un caso

9. Incidencia de cada categoría de errores en el corpus de análisis

en que la palabra se dejó sin traducir. También se añadieron 2 elementos tipográficos y se trató de traducir erróneamente una URL en una ocasión.

En la ilustración a continuación presentamos gráficamente la distribución de las ediciones realizadas. Aquí se puede ver claramente que la mayoría de las ediciones se hicieron en la corrección de la traducción de nombres propios:

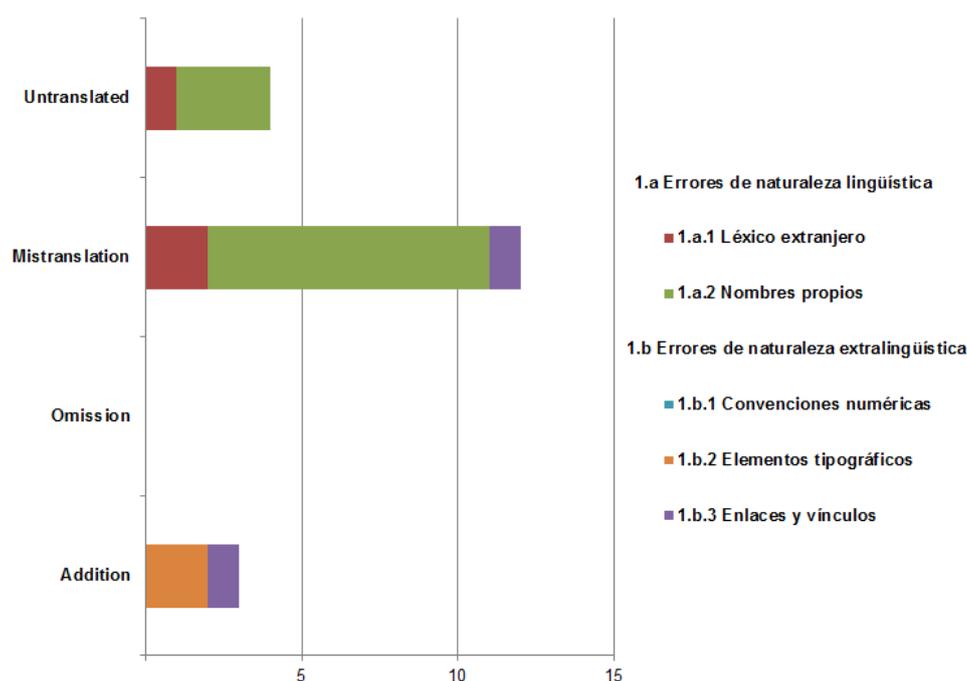


Ilustración 91: Relación del motor SMT1 en cada categoría de la primera macrocategoría

En la tabla 203 podemos comprobar que las acciones más habituales en este motor y categoría se dan para editar *mistranslations* y *untranslated* nombres propios y extranjerismos. Ahora veamos si en el otro motor de este tipo cambian los errores editados:

smt2	Categoría	Valor del Error					Total Categoría				
		Addition		Omission		Mistranslation		Untranslated		Misplaced	
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
1	Errores independientes de la combinación lingüística										
1.a	Errores de naturaleza lingüística										
1.a.1	Léxico extranjero					5	(5,21)	1	(1,04)		
1.a.2	Nombres propios					2	(2,08)	6	(6,25)		
1.b	Errores de naturaleza extralingüística										
1.b.1	Convenciones numéricas										
1.b.2	Elementos tipográficos	4	(4,17)							1	(1,04)
1.b.3	Enlaces y vínculos										

Tabla 205: Número y porcentaje total de errores marcados en la primera macrocategoría en el motor SMT2

9. Incidencia de cada categoría de errores en el corpus de análisis

También en este motor el error más frecuente es la traducción incorrecta de nombres propios. En dos casos se tradujeron erróneamente y en seis casos se dejaron de traducir. En cuanto a las palabras en léxico extranjero se marcaron cinco casos de *mistranslation* y un caso en que la palabra se dejó sin traducir. También se añadieron 4 elementos tipográficos y en una ocasión se colocó un elemento en una posición incorrecta.

En la ilustración a continuación presentamos gráficamente la distribución de las ediciones realizadas. En esta gráfica queda claro que en las propuestas de este motor no se editaron errores en la traducción o localización de vínculos y que se realizaron más ediciones por errores en la traducción de léxico extranjero que con el motor SMT1:

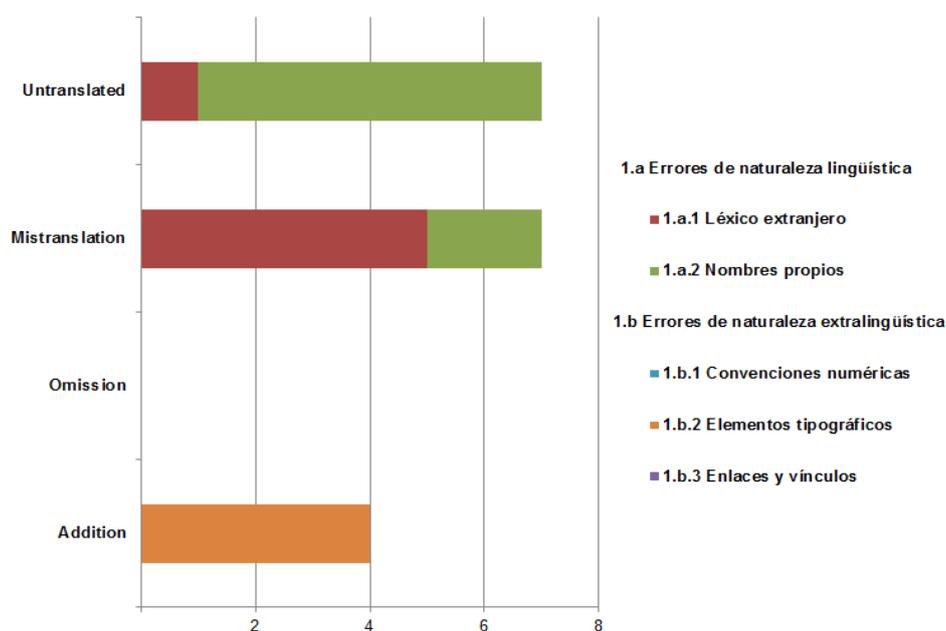


Ilustración 92: Relación del motor SMT2 en cada categoría de la primera macrocategoría

Vemos que la incidencia de *mistranslations* y *untranslated* entre nombres propios y extranjerismos es aún más marcada que en el caso del SMT1.

Ahora veamos qué sucede cuando las propuestas seleccionadas provienen de motores basados en reglas, empezaremos con la tabla de descripción de las propuestas del RBMT1:

9. Incidencia de cada categoría de errores en el corpus de análisis

rbmt1	Indicador	Categoría	Valor del Error					Total Categoría				
			Addition		Omission		Mistranslation		Untranslated		Misplaced	
			Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
1		Errores independientes de la combinación lingüística										
1.a		Errores de naturaleza lingüística										
1.a.1		Léxico extranjero				5 (5,21)		8 (8,33)				13
1.a.2		Nombres propios				16 (16,67)		12 (12,50)				28
1.b		Errores de naturaleza extralingüística										
1.b.1		Convenciones numéricas										
1.b.2		Elementos tipográficos	1	(1,04)						1	(1,04)	2
1.b.3		Enlaces y vínculos										

Tabla 206: Número y porcentaje total de errores marcados en la primera macrocategoría en el motor RBMT1

En el error RBMT1 el error más frecuente es la traducción incorrecta de nombres propios. En 16 casos se tradujeron erróneamente y en 12 casos se dejaron de traducir. En cuanto a las palabras en léxico extranjero se marcaron cinco casos de *mistranslation* y ocho caso en que las palabras se dejaron sin traducir. Aquí a diferencia del motor anterior solo se añadió un elemento tipográfico y en una ocasión se colocó uno de estos elementos en una posición incorrecta.

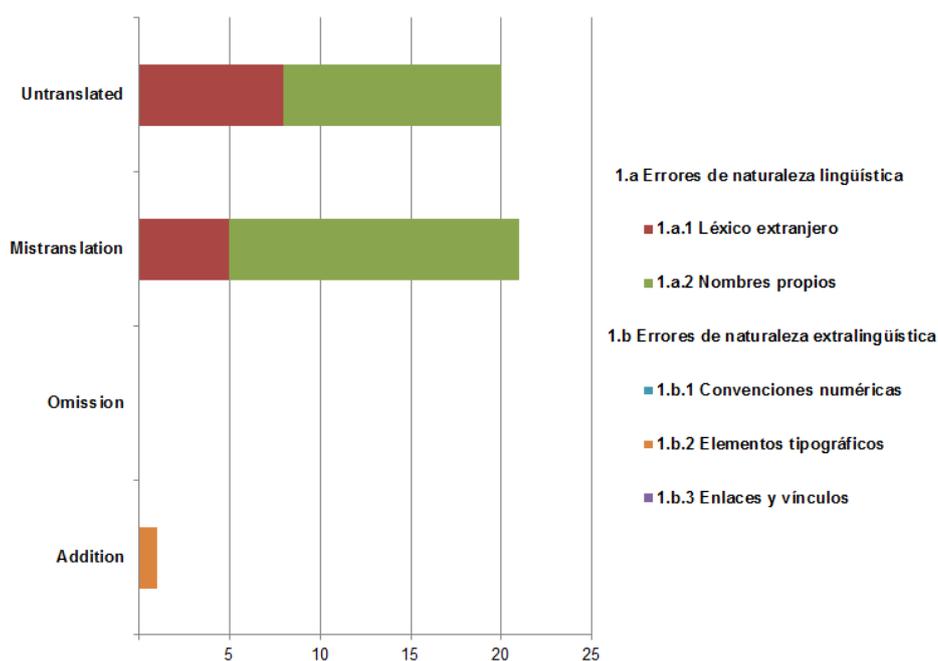


Ilustración 93: Relación del motor RBMT1 en cada categoría de la primera macrocategoría

En la ilustración se puede ver que no se realizó ninguna edición que tuviese que ver con omisión de palabra, nombres propios, elementos tipográficos o enlaces. En el RBMT2 observamos lo siguiente:

9. Incidencia de cada categoría de errores en el corpus de análisis

rbmt2	Indicador	Categoría	Valor del Error						Total Categoría				
			Addition		Omission		Mistranslation			Untranslated		Misplaced	
			Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
1		Errores independientes de la combinación lingüística											
1.a		Errores de naturaleza lingüística											
1.a.1		Léxico extranjero						3	(3,13)				3
1.a.2		Nombres propios				5	(5,21)						5
1.b		Errores de naturaleza extralingüística											
1.b.1		Convenciones numéricas											
1.b.2		Elementos tipográficos			2	(2,08)					5	(5,21)	7
1.b.3		Enlaces y vínculos											

Tabla 207: Número y porcentaje total de errores marcados en la primera macrocategoría en el motor RBMT2

Como en los otros motores el error más frecuente es la traducción incorrecta de nombres propios. En 5 casos se tradujeron erróneamente. En cuando a las palabras en léxico extranjero se marcaron tres casos en que las palabras se dejaron sin traducir. Aquí a diferencia del motor anterior se omitieron dos elementos tipográficos y en cinco casos se colocaron en una posición incorrecta.

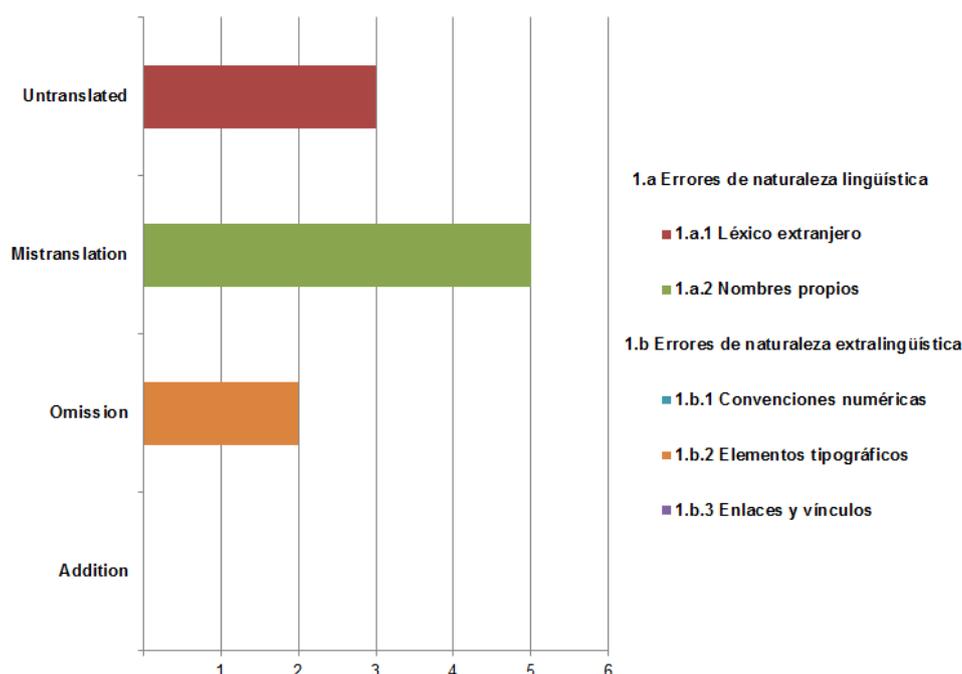


Ilustración 94: Relación del motor RBMT2 en cada categoría de la primera macrocategoría

En la ilustración vemos la relación ponderada de los resultados de las ediciones de todos los motores en esta categoría. Hemos ponderado los resultados porque el número de propuestas seleccionadas por motor es diferente (ver página 232):

9. Incidencia de cada categoría de errores en el corpus de análisis

Si comparamos los resultados, la categoría en que se realizó un mayor número de ediciones en esta macrocategoría en general fue la que utilizamos para marcar los errores en la traducción o tratamiento de nombres propios (53 ediciones); en segundo lugar se realizaron más ediciones por errores en el tratamiento de léxico extranjero (25 casos); no se presentó en nuestro corpus ningún errores en el tratamiento de convenciones numéricas, pero si presentaron errores en la interpretación de elementos tipográficos (16 ediciones) y de enlaces y vínculos (2 ediciones). La distribución en cada motor tuvo un elemento en común. En todos los motores se realizaron un mayor número de ediciones en la corrección de la traducción o tratamiento erróneo de nombres propios:

Motor	1.a.1 Léxico extranjero	1.a.2 Nombres propios	1.b.1 Convenciones	1.b.2 Elementos tipográficos	1.b.3 Enlaces (URLs)
SMT1	3	12		2	2
SMT2	6	8			5
RBMT1	13	28		2	
RBMT2	3	5		7	

Tabla 208: Categorías con más número de ediciones por motor

Como estrategia para la posesición podríamos establecer que se debe tener en cuenta que los motores de traducción automática no pueden reconocer siempre cuando una palabra es o no un nombre propio y cuándo debe o no traducirse. Veamos ahora si en las categorías de la segunda macrocategoría también hay una categoría en la que se hayan realizado más ediciones en todos los motores.

9.3.2 Segunda macrocategoría errores provocados por las características propias del alemán

En las categorías que se agrupan bajo esta macrocategoría las diferencias entre los tipos de motores no son en la mayoría de los casos tan considerables.

9. Incidencia de cada categoría de errores en el corpus de análisis

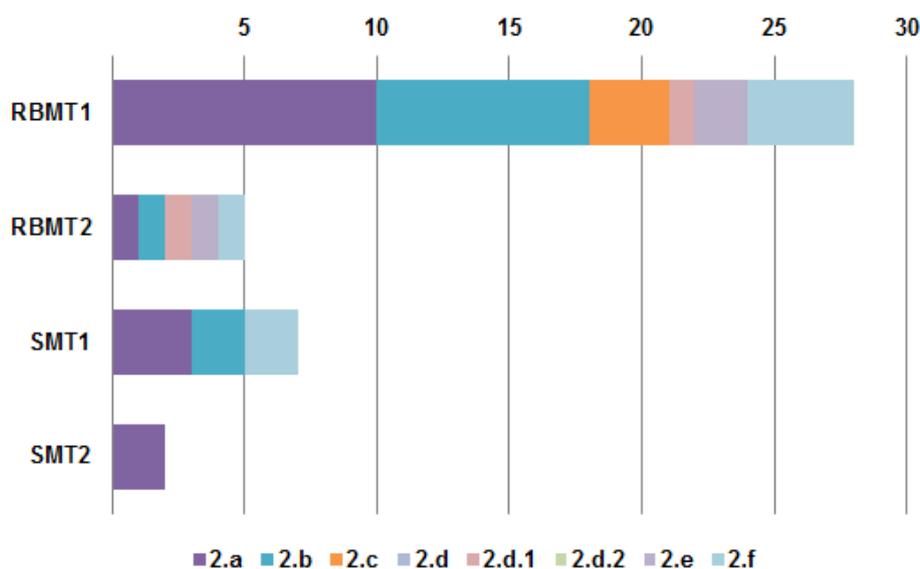


Ilustración 95: Relación tipo de motor categorías de la segunda macrocategoría

Hay tres puntos a destacar. En los motores basados en reglas no se omitieron las traducciones de los verbos separables. En los motores estadísticos sí. Tampoco encontramos muchos problemas en la transferencia sintáctica de composiciones, pero ello también puede darse porque la mayoría de las propuestas seleccionadas eran propuestas del motor de traducción RBMT1. Del mismo modo, en las propuestas de los RBMT2 aparecen errores en la reproducción de la negación, mientras que en los SMT no hay errores de este tipo, porque las propuestas erróneas que pudiera haber no fueron consideradas para la posesición.

Veamos los detalles. Empezaremos con el motor SMT1. Primero la tabla detallada de conteo de errores:

smt1	Indicador	Categoría	Valor del Error						Total Categoría
			Addition		Omission		Mistranslation		
			Núm.	(%)	Núm.	(%)	Núm.	(%)	
2.a	Verbos separables			2	(4,76)	1	(2,38)	3	
2.b	Conjunciones y partículas					2	(4,76)	2	
2.c	Composiciones								
2.d	Oraciones compuestas:								
	2.d.1 Oraciones subordinadas								
	2.d.2 Oraciones coordinadas								
2.e	Negación								
2.f	Colocaciones			1	(2,38)	1	(2,38)	2	

Tabla 209: Número y porcentaje total de errores marcados en la segunda macrocategoría en el motor SMT1

9. Incidencia de cada categoría de errores en el corpus de análisis

Aquí podemos observar que en esta categoría se realizaron pocas ediciones, la mayoría tiene que ver con problemas en la traducción de verbos separables: en 2 casos se omitieron y en 1 caso se tradujo correctamente el verbo. También se omitió una colocación o fraseología y en otro caso no se tradujo correctamente.

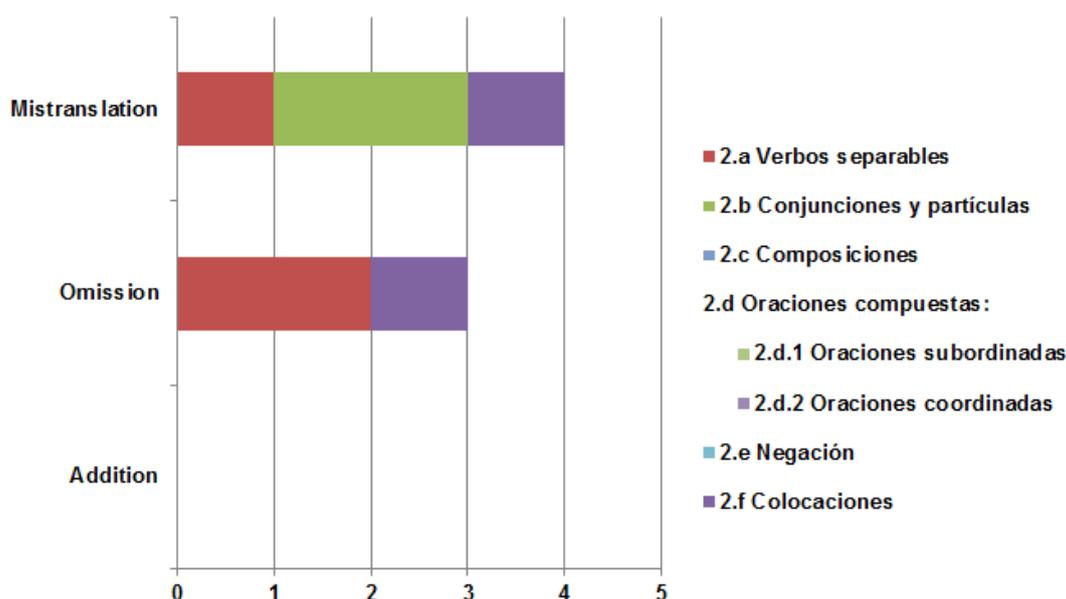


Ilustración 96: Relación del motor SMT1 en cada categoría de la segunda macrocategoría

En la ilustración se ve claramente que no se presentaron adiciones; no incluimos en la ilustración el valor *untranslated* porque tampoco se realizan ediciones por este motivo. Veamos cómo es la relación de ediciones marcadas en el SMT2. Aquí tenemos solo dos casos de traducción errónea de verbos separables.

smt2	Indicador	Categoría	Valor del Error						Total Categoría
			Addition		Omission		Mistranslation		
			Núm.	(%)	Núm.	(%)	Núm.	(%)	
	2.a	Verbos separables					2	(4,76)	2
	2.b	Conjunciones y partículas							
	2.c	Composiciones							
	2.d	Oraciones compuestas:							
	2.d.1	Oraciones subordinadas							
	2.d.2	Oraciones coordinadas							
	2.e	Negación							
	2.f	Colocaciones							

Tabla 210: Número y porcentaje total de errores marcados en la segunda macrocategoría en el motor SMT2

9. Incidencia de cada categoría de errores en el corpus de análisis

Como solo se dio este error, no consideramos necesario añadir una ilustración. En los motores basados en reglas encontramos la siguiente relación. Como se seleccionaron más propuestas del motor RBMT1, por ello tenemos aquí más información.

rbmt1		Valor del Error								Total Categoría
		Addition		Omission		Mistranslation		Untranslated		
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
2.a	Verbos separables					10	(23,81)			10
2.b	Conjunciones y partículas					7	(16,67)	1	(2,38)	8
2.c	Composiciones			1	(2,38)	2	(4,76)			3
2.d	Oraciones compuestas:									
	2.d.1 Oraciones subordinadas					1	(2,38)			1
	2.d.2 Oraciones coordinadas									
2.e	Negación					2	(4,76)			2
2.f	Colocaciones					4	(9,52)			4

Tabla 211: Número y porcentaje total de errores marcados en la segunda macrocategoría en el motor RBMT1

En el motor RBMT1 se editaron 10 *mistranslation* de verbos separables. 8 ediciones en la traducción de conjunciones y partículas, 3 ediciones en la traducción de composiciones (1 por omisión y 2 por *mistranslation*). Se realizaron 2 ediciones porque no se reprodujo correctamente el sentido de la negación y 4 por *mistranslation* de colocaciones o fraseologías. Solo se editó un error en una oración compuesta.

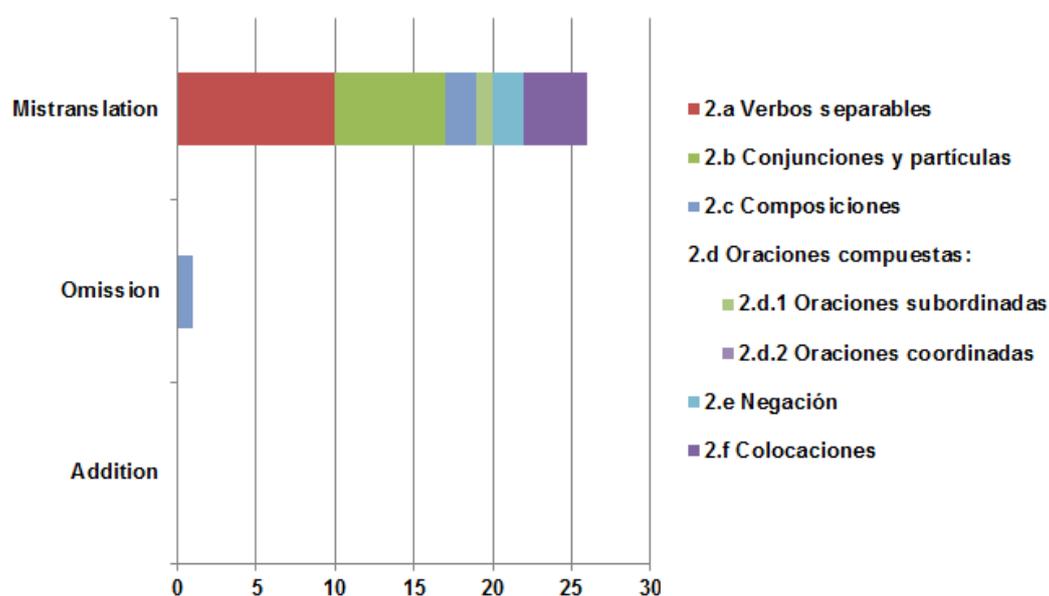


Ilustración 97: Relación del motor RBMT1 en cada categoría de la segunda macrocategoría

9. Incidencia de cada categoría de errores en el corpus de análisis

Las ediciones en el RBMT2 son del siguiente orden, 1 *mistranslation* de un verbo separable, 1 edición en la traducción de conjunciones y partículas, 3 ediciones en la traducción de composiciones (1 por omisión y 2 por *mistranslation*). También se marcó un error respectivamente porque no se reprodujo correctamente el sentido de la negación y por *mistranslation* de colocaciones o fraseologías y por un error en una oración compuesta.

rbmt2		Valor del Error								Total Categoría
		Addition		Omission		Mistranslation		Untranslated		
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
2.a	Verbos separables					1	(2,38)			1
2.b	Conjunciones y partículas					1	(2,38)			1
2.c	Composiciones									
2.d	Oraciones compuestas:									
2.d.1	Oraciones subordinadas					1	(2,38)			1
2.d.2	Oraciones coordinadas									
2.e	Negación					1	(2,38)			1
2.f	Colocaciones					1	(2,38)			1

Tabla 212: Número y porcentaje total de errores marcados en la segunda macrocategoría en el motor RBMT2

La categoría en que se realizó un mayor número de ediciones en esta macrocategoría en general fue la que utilizamos para marcar las ediciones de errores en la traducción de verbos separables (16 ediciones); en segundo lugar se realizaron más ediciones por errores en la traducción de conjunciones o partículas (11 casos) y en tercer lugar en la traducción de colocaciones (7 casos); solo tendremos en cuenta para la comparación de los resultados en los motores las categorías de esta macrocategoría en que se presentaron más errores:

Motor	2.a Verbos separables	2.b Conjunciones	2.c Composiciones	2.f Colocaciones
SMT1	3	2		2
SMT2	2			0
RBMT1	10	8	3	4
RBMT2	1	1	0	1

Tabla 213: Categorías con más número de ediciones por motor

Como podemos ver en la tabla anterior, la categoría con más errores es en la que marcamos las ediciones en la traducción de verbos separables. A pesar de que los números de ediciones no son tantos como en las otras categorías hay que considerar que en nuestra combinación lingüística la traducción de los verbos separables debe ser tomada en cuenta para la

9. Incidencia de cada categoría de errores en el corpus de análisis

posedición, ya que se pueden dar muchos errores si estos verbos no se interpretan correctamente. Continuemos el análisis con la siguiente categoría.

9.3.3 Tercera macrocategoría: errores provocados por características del castellano

El análisis de los errores marcados en esta macrocategoría es más complejo porque esta agrupa más categorías que las macrocategorías anteriores. Aquí podemos ver que se marcaron más ediciones en la flexión verbal en las propuestas de los RBMT que en las de los motores estadísticos. Mientras que los motores estadísticos presentan más problemas en lo que tiene que ver con la concordancia de número y el orden sintáctico.

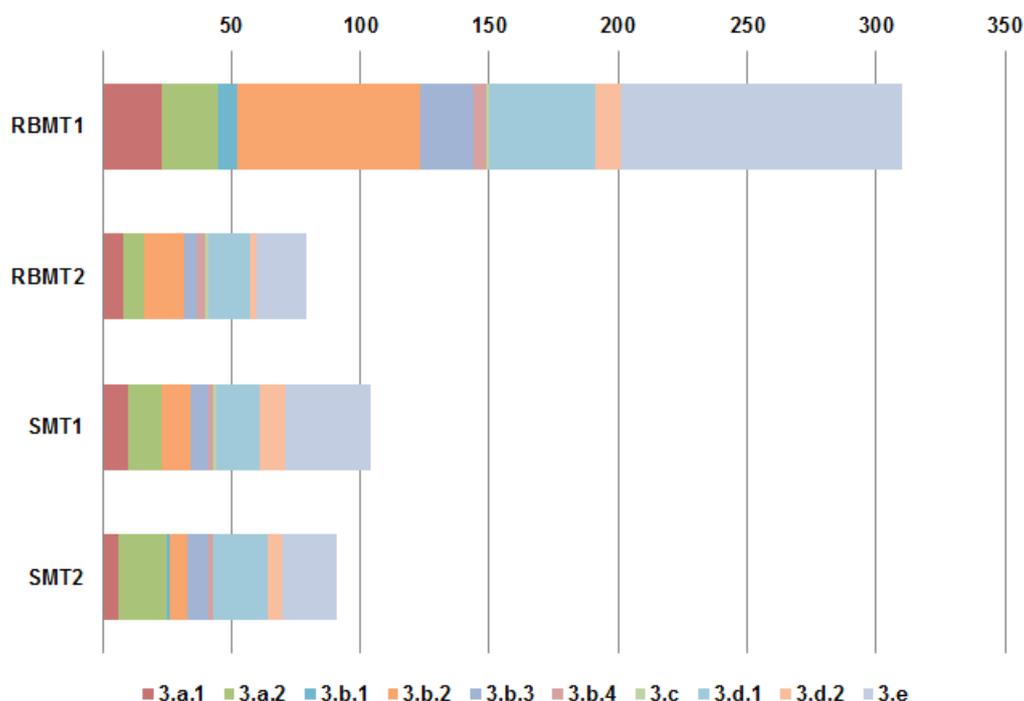


Ilustración 98: Relación por motor y categoría en la tercera macrocategoría

Ahora veamos cómo es esta relación en cada motor. Empezaremos como en los apartados anteriores con el motor SMT1. Primero la tabla detallada de conteo de errores:

9. Incidencia de cada categoría de errores en el corpus de análisis

smt1		Valor del Error										Total Categoría				
		Addition		Omission		Mistranslation		Questionable		Misplaced			Register		Not idiomatic	
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
3.a	Errores de flexión o concordancia															
3.a.1	Concordancia de género					10	(1,71)									
3.a.2	Concordancia de número					13	(2,23)									
3.b	Errores de flexión verbal															
3.b.1	Flexión verbal (tiempo)															
3.b.2	Flexión verbal (modo)					11	(1,88)									
3.b.3	Flexión verbal (número)					7	(1,20)									
3.b.4	Flexión verbal (voz)					2	(0,34)									
3.c	Estilo													1	(0,17)	
3.d	Errores ortotipográficos															
3.d.1	Ortografía	7	(1,20)	7	(1,20)			3	(0,51)							
3.d.2	Tipografía	1	(0,17)	8	(1,37)					1	(0,17)					
3.e	Sintaxis	3	(0,51)							30	(5,14)					
																33

Tabla 214: Número y porcentaje total de errores marcados en la tercera macrocategoría en el motor SMT1

Aquí podemos observar que el mayor número de ediciones tiene que ver con problemas de orden sintáctico (33 casos), concordancia de género o número (23 casos) y de flexión verbal (20 errores). Tenemos 43 casos de *mistranslation*, 11 casos de omisión, 15 casos de adición y 31 casos de *misplaced*, que se dan en la sintaxis.

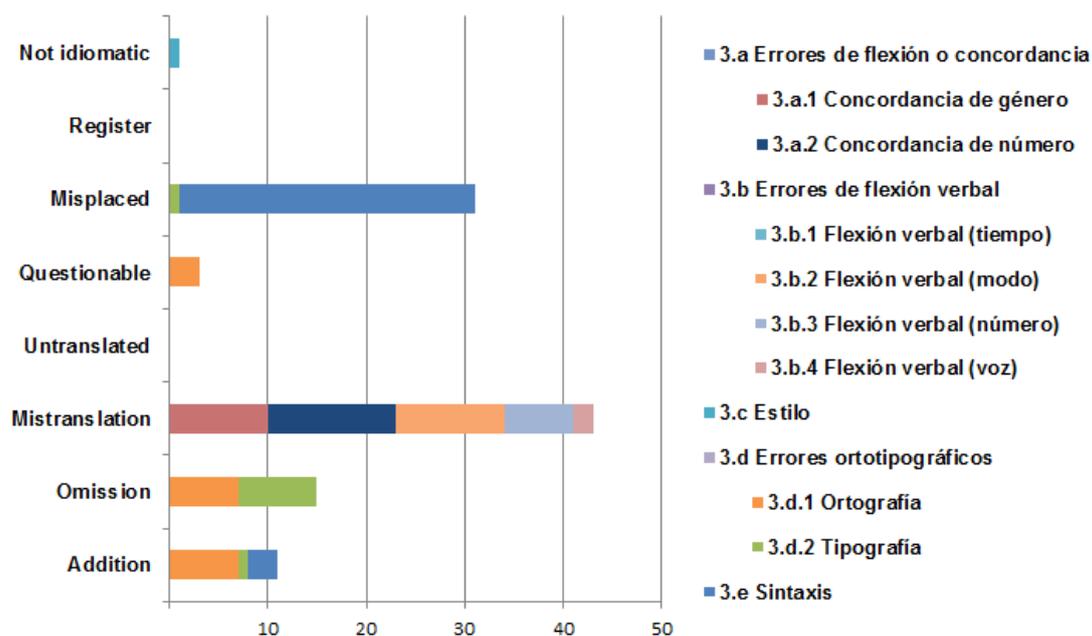


Ilustración 99: Relación del motor SMT1 en cada categoría de la tercera macrocategoría

Ahora veamos cómo es la relación de ediciones marcadas en el SMT2. Aquí tenemos 43 ediciones por *mistranslation*, 12 por adición, 10 por omisión, 22 por *misplaced* que en este motor tiene que ver tanto con la sintaxis como con errores ortotipográficos.

9. Incidencia de cada categoría de errores en el corpus de análisis

smt2	Indicador	Categoría	Valor del Error										Total Categoría				
			Addition		Omission		Mistranslation		Questionable		Misplaced			Register		Not idiomatic	
			Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
	3.a	Errores de flexión o concordancia															
	3.a.1	Concordancia de género					6	(1,03)									6
	3.a.2	Concordancia de número					19	(3,25)									19
	3.b	Errores de flexión verbal															
	3.b.1	Flexión verbal (tiempo)					1	(0,17)									1
	3.b.2	Flexión verbal (modo)					7	(1,20)									7
	3.b.3	Flexión verbal (número)					8	(1,37)									8
	3.b.4	Flexión verbal (voz)					2	(0,34)									2
	3.c	Estilo															
	3.d	Errores ortotipográficos															
	3.d.1	Ortografía	12	(2,05)	4	(0,68)			3	(0,51)	2	(0,34)					21
	3.d.2	Tipografía			4	(0,68)			1	(0,17)	1	(0,17)					6
	3.e	Sintaxis			2	(0,34)					19	(3,25)					21

Tabla 215: Número y porcentaje total de errores marcados en la tercera macrocategoría en el motor SMT2

En las propuestas del motor SMT2 seleccionadas se marcaron 25 errores de flexión de género o número, 21 errores de orden sintáctico y 18 errores de flexión verbal. Es decir, se corrigen el mismo tipo de errores, en una proporción ligeramente diferente. En la siguiente ilustración podemos ver la distribución:

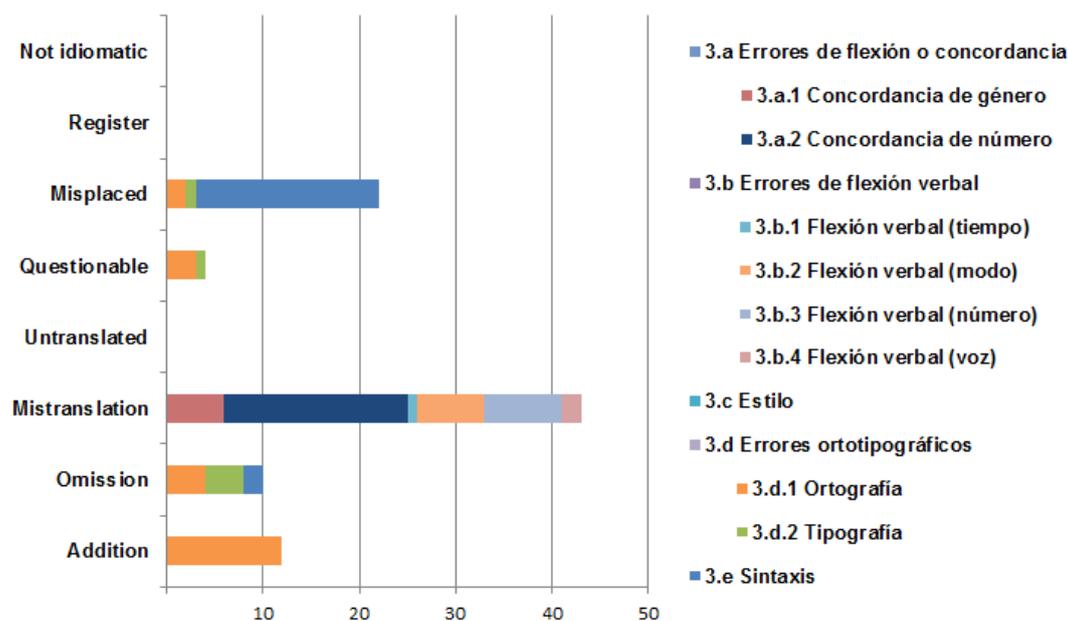


Ilustración 100: Relación del motor SMT2 en cada categoría de la tercera macrocategoría

En el RBMT1 encontramos la siguiente relación de número y el porcentaje de los valores respectivos en las categorías de errores que tienen que ver con errores provocados por características propias del castellano:

9. Incidencia de cada categoría de errores en el corpus de análisis

rbmt1		Valor del Error										Total Categoría				
		Addition		Omission		Mistranslation		Questionable		Misplaced			Register		Not idiomatic	
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
3.a	Errores de flexión o concordancia															
3.a.1	Concordancia de género					23	(3,94)									23
3.a.2	Concordancia de número					22	(3,77)									22
3.b	Errores de flexión verbal															
3.b.1	Flexión verbal (tiempo)					7	(1,20)									7
3.b.2	Flexión verbal (modo)					71	(12,16)									71
3.b.3	Flexión verbal (número)					21	(3,60)									21
3.b.4	Flexión verbal (voz)					5	(0,86)									5
3.c	Estilo													1	(0,17)	1
3.d	Errores ortotipográficos															
3.d.1	Ortografía	10	(1,71)	8	(1,37)			21	(3,60)	2	(0,34)					41
3.d.2	Tipografía	1	(0,17)	8	(1,37)					1	(0,17)					10
3.e	Sintaxis			1	(0,17)					108	(18,49)					109

Tabla 216: Número y porcentaje total de errores marcados en la tercera macrocategoría en el motor RBMT1

En el motor RBMT1 se editaron 149 *mistranslation*, 16 omisiones y 11 adiciones. Se realizaron 109 ediciones por errores de sintaxis y 5 correcciones de ortotipografía. De los 149 casos marcados como *mistranslation* la mayoría (104) son errores de flexión verbal y los 45 casos restantes errores de flexión o concordancia de género o número.

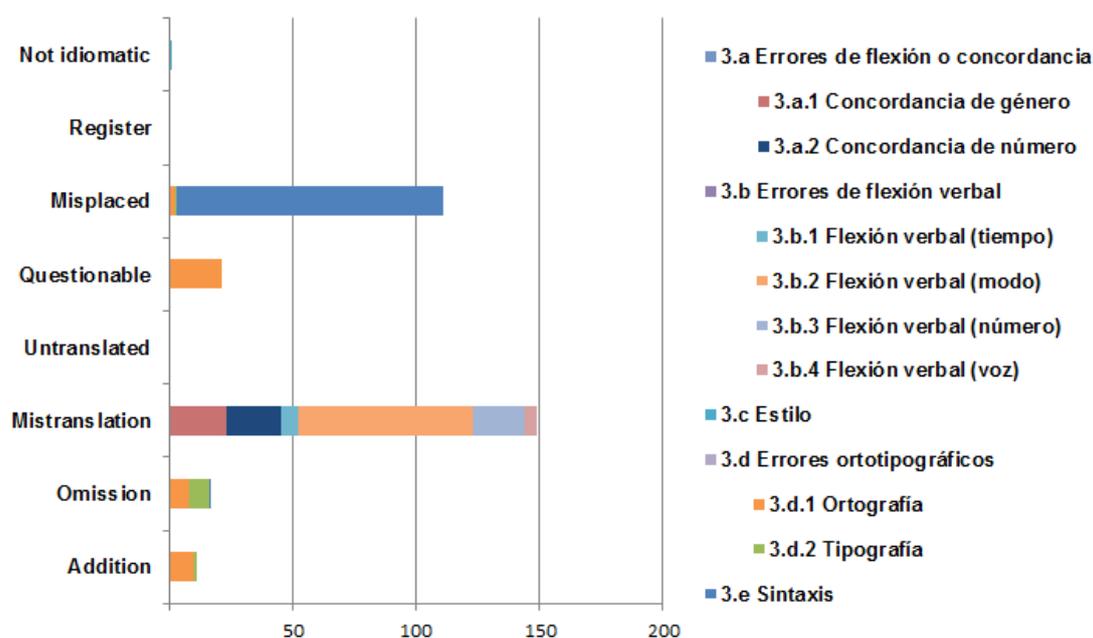


Ilustración 101: Relación del motor RBMT1 en cada categoría de la tercera macrocategoría

Las ediciones en el RBMT2 son del siguiente orden: en 40 casos se corrigieron errores por *mistranslation*, 6 adiciones y 2 omisiones. También encontramos 21 *misplaced*, 19 de los cuales son errores de orden sintáctico. En las propuestas de este motor encontramos un caso en que se editó un error de registro.

9. Incidencia de cada categoría de errores en el corpus de análisis

rbmt2		Valor del Error										Total Categoría				
		Addition		Omission		Mistranslation		Questionable		Misplaced			Register		Not idiomatic	
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
3.a	Errores de flexión o concordancia															
3.a.1	Concordancia de género					8	(1,37)									8
3.a.2	Concordancia de número					8	(1,37)									8
3.b	Errores de flexión verbal															
3.b.1	Flexión verbal (tiempo)															
3.b.2	Flexión verbal (modo)					16	(2,74)									16
3.b.3	Flexión verbal (número)					5	(0,86)									5
3.b.4	Flexión verbal (voz)					3	(0,51)									3
3.c	Estilo											1	(0,17)			1
3.d	Errores ortotipográficos															
3.d.1	Ortografía	6	(1,03)					9	(1,54)	1	(0,17)					16
3.d.2	Tipografía			2	(0,34)					1	(0,17)					3
3.e	Sintaxis									19	(3,25)					19

Tabla 217: Número y porcentaje total de errores marcados en la tercera macrocategoría en el motor RBMT2

En el motor RBMT2 se editaron solo 40 *mistranslation*, —menos que en el motor RBMT1— 2 omisiones y 6 adiciones. Se realizaron 19 ediciones por errores de sintaxis, 2 correcciones de ortotipografía. De los 40 casos marcados como *mistranslation* 24 son errores de flexión verbal y los 16 casos restantes errores de flexión o concordancia de género o número.

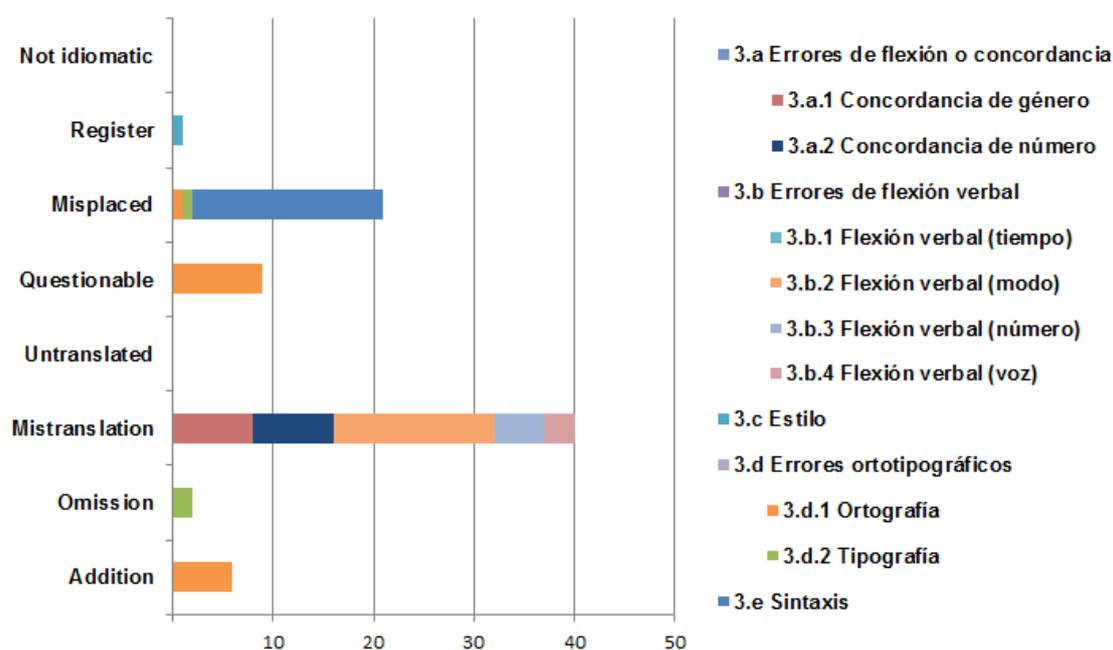


Ilustración 102: Relación del motor RBMT2 en cada categoría de la tercera macrocategoría

La categoría en que se realizó un mayor número de ediciones en esta macrocategoría en general fue la que utilizamos para marcar las ediciones de errores sintaxis (182 ediciones); en segundo lugar, se realizaron más ediciones por errores en la conjugación (166 casos) y por último por errores de concordancia de número o género (109 casos); sin embargo la distribución en los motores del mismo tipo fue ligeramente diferente:

9. Incidencia de cada categoría de errores en el corpus de análisis

Motor	3.a Concordancia	3.b Flexión verbal	3.d Sintaxis
SMT1	13	20	33
SMT2	25	18	21
RBMT1	45	104	109
RBMT2	16	24	19

Tabla 218: Categorías con más número de ediciones por motor

Como podemos ver en la tabla anterior, la categoría con más errores cambia en los motores del mismo tipo. Y esto se repite con las categorías en segundo y tercer lugar. Veamos ahora en la cuarta macrocategoría si este fenómeno se repite.

9.3.4 Cuarta macrocategoría: errores en la microestructura

Esta macrocategoría agrupa dos categorías principales (palabras con y sin carga semántica), y aquí también analizaremos los errores encontrados en las propuestas de los motores teniendo en cuenta esta división. Así que empezaremos con el análisis de las ediciones realizadas que tienen que ver con palabras sin carga semántica, como preposiciones, determinantes, adverbios y conjunciones.

9.3.4.1 Palabras sin carga semántica

En la ilustración a continuación vemos la distribución general de errores en esta subcategoría y teniendo en cuenta cada motor. Como el número de propuestas seleccionadas varía en cada motor, hicimos el cálculo ponderado.

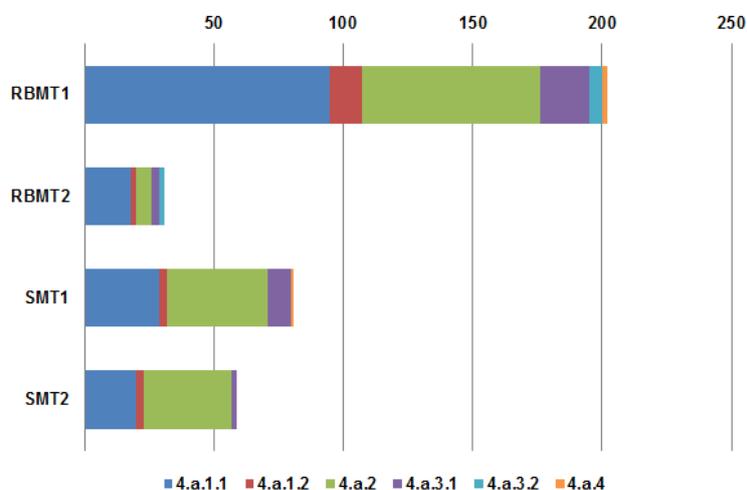


Ilustración 103: Relación tipo de motor categorías de la cuarta macrocategoría sin carga lingüística

9. Incidencia de cada categoría de errores en el corpus de análisis

Ahora veamos cómo es esta relación en cada motor. Empezaremos como en el apartado anterior con el motor SMT1. Primero la tabla detallada de conteo de errores:

smt1	Indicador	Categoría	Valor del Error						Total Categoría
			Addition		Omission		Mistranslation		
			Núm.	(%)	Núm.	(%)	Núm.	(%)	
4.a	Palabras sin carga semántica								
4.a.1	Errores de carácter preposicional								
4.a.1.1	Preposiciones		5	(0,47)	8	(0,75)	16	(1,51)	29
4.a.1.2	Estructuras preposicionales						3	(0,28)	3
4.a.2	Determinantes		23	(2,16)	11	(1,03)	5	(0,47)	39
4.a.3	Errores de carácter adverbial								
4.a.3.1	Adverbios		2	(0,19)	2	(0,19)	5	(0,47)	9
4.a.3.2	Estructuras adverbiales								
4.a.4	Conjunciones y partículas		1	(0,09)					1

Tabla 219: Número y porcentaje total de errores en la cuarta macrocategoría (sin carga semántica) en el motor SMT1

Aquí podemos observar que el mayor número de ediciones se dan en relación con la traducción de determinantes (39 casos), la traducción de preposiciones (31 casos) y de adverbios (9 errores). Tenemos 29 casos de *mistranslation*, 21 casos de omisión y 31 casos de adición; los otros valores no fueron relevantes para esta categoría. A diferencia de la macrocategoría anterior, se dan aquí muchos más casos de omisión o adición.

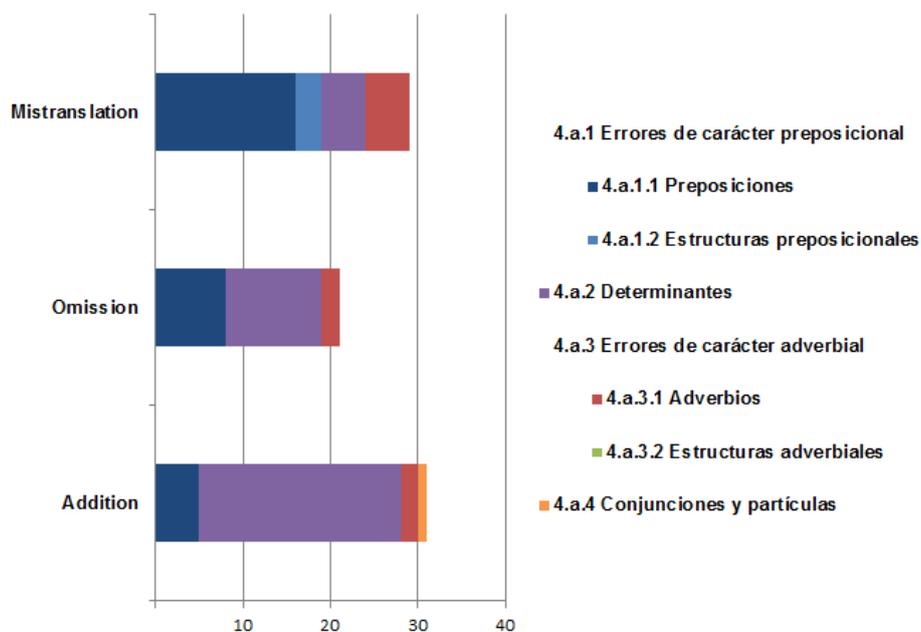


Ilustración 104: Relación del motor SMT1 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica

9. Incidencia de cada categoría de errores en el corpus de análisis

Ahora veamos cómo es la relación de ediciones marcadas en el SMT2. Podemos observar que el mayor número de ediciones también se da en relación con la traducción de determinantes (34 casos), la traducción de preposiciones y estructura adverbial (23 casos) y solo se presentan 2 errores en la traducción de adverbios. Tenemos 15 casos de *mistranslation*, 17 casos de omisión y 27 casos de adición; los otros valores no fueron tampoco con este motor relevantes para esta categoría. A diferencia de la macrocategoría anterior, en el SMT2 también se dan aquí muchos más casos de omisión o adición.

smt2		Valor del Error						Total Categoría
		Addition		Omission		Mistranslation		
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	
4.a	Palabras sin carga semántica							
4.a.1	Errores de carácter preposicional							
4.a.1.1	Preposiciones	4	(0,38)	6	(0,56)	10	(0,94)	20
4.a.1.2	Estructuras preposicionales					3	(0,28)	3
4.a.2	Determinantes	22	(2,07)	10	(0,94)	2	(0,19)	34
4.a.3	Errores de carácter adverbial							
4.a.3.1	Adverbios	1	(0,09)	1	(0,09)			2
4.a.3.2	Estructuras adverbiales							
4.a.4	Conjunciones y partículas							

Tabla 220: Número y porcentaje total de errores en la cuarta categoría de palabras sin carga semántica en el motor SMT2

Como en las propuestas del motor SMT1, en las propuestas seleccionadas del motor SMT2 la categoría en la que se marcaron más ediciones que ver con la traducción de determinantes; en este motor no se marcó ninguna edición en la traducción de conjunciones o particulares. En la siguiente ilustración podemos ver la distribución:

9. Incidencia de cada categoría de errores en el corpus de análisis

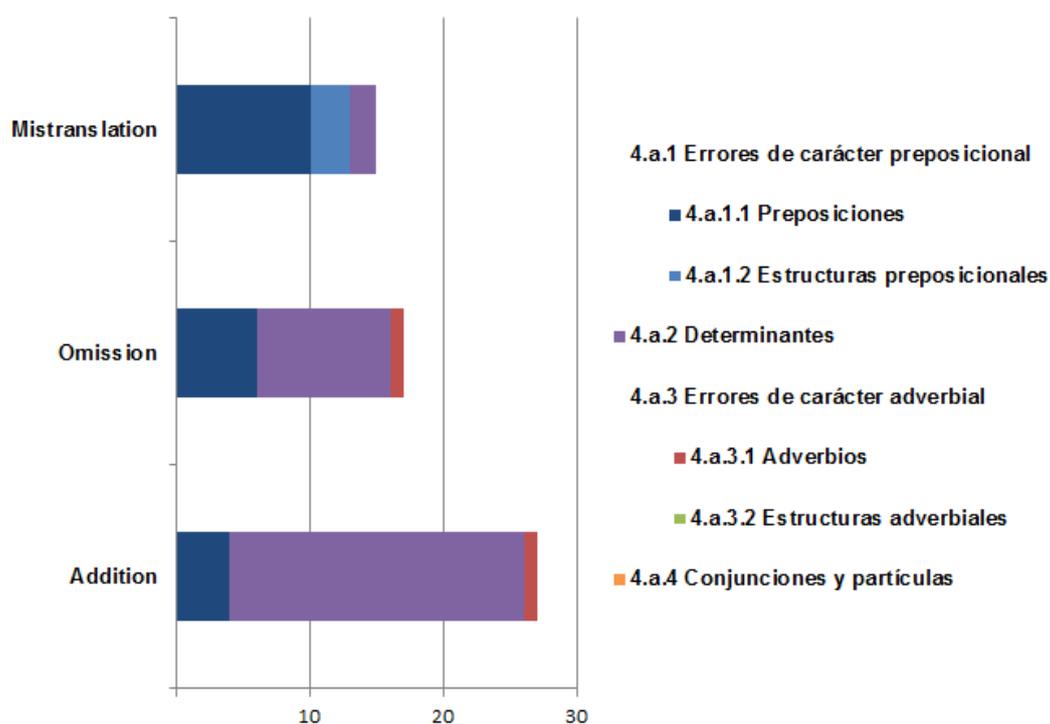


Ilustración 105: Relación del motor SMT2 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica

Veamos que sucede con los motores basados en reglas, como en el apartado anterior empezaremos con el motor RBMT1.

rbmt1	Indicador	Categoría	Valor del Error						Total Categoría
			Addition		Omission		Mistranslation		
			Núm.	(%)	Núm.	(%)	Núm.	(%)	
4.a		Palabras sin carga semántica							
	4.a.1	Errores de carácter preposicional							
	4.a.1.1	Preposiciones	21	(1,98)	19	(1,79)	54	(5,08)	95
	4.a.1.2	Estructuras preposicionales	1	(0,09)	1	(0,09)	10	(0,94)	12
	4.a.2	Determinantes	19	(1,79)	45	(4,23)	5	(0,47)	69
	4.a.3	Errores de carácter adverbial							
	4.a.3.1	Adverbios	3	(0,28)			16	(1,51)	19
	4.a.3.2	Estructuras adverbiales					5	(0,47)	5
	4.a.4	Conjunciones y partículas	1	(0,09)	1	(0,09)			2

Tabla 221: Número y porcentaje total de errores en la cuarta categoría de palabras sin carga semántica en el motor RBMT1

En el motor RBMT1 se editaron 90 *mistranslations*, 66 omisiones y 45 adiciones. También encontramos un caso en que se editó una preposición que no se tradujo. Aquí encontramos una diferencia en relación con los motores estadísticos; las ediciones que se realizaron tienen que ver con las preposiciones y no —como en los motores estadísticos— con los de-

9. Incidencia de cada categoría de errores en el corpus de análisis

terminantes. Se marcaron 95 ediciones de preposiciones y 12 de estructuras preposicionales. En segundo lugar, se marcaron 69 errores en la traducción de determinantes y en tercer lugar 24 errores en la traducción de adverbios y estructuras adverbiales. En la ilustración a continuación se puede ver la distribución de los errores:

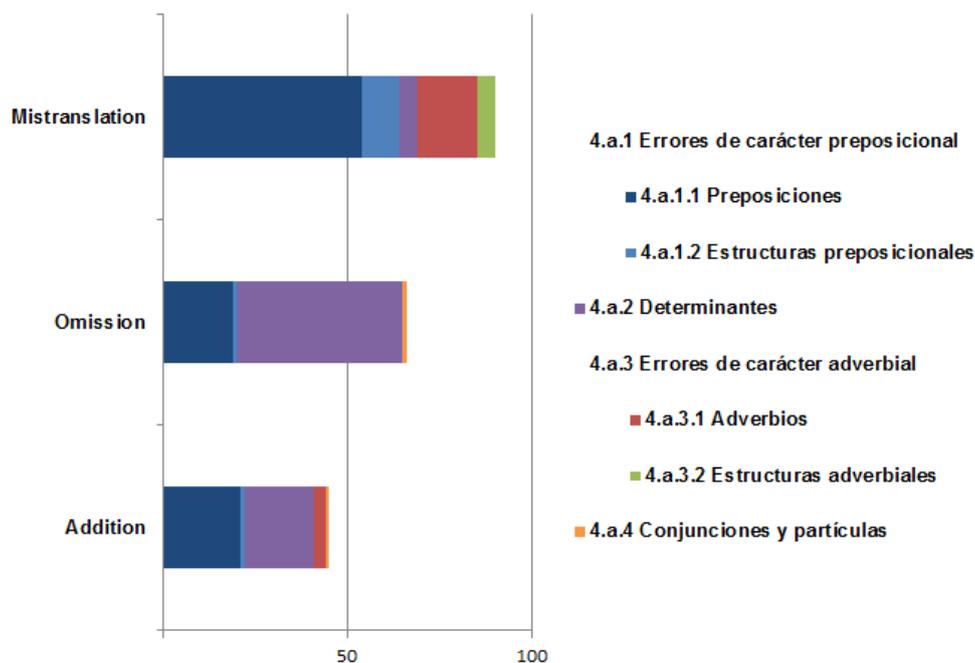


Ilustración 106: Relación del motor RBMT1 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica

Las ediciones en el RBMT2 son del siguiente orden, en 12 casos se editaron errores por *mistranslation*, 5 adiciones y 14 omisiones. También encontramos 21 *misplaced*, 19 de los cuales son errores de orden sintáctico.

rbmt2	Categoría	Valor del Error						Total Categoría
		Addition		Omission		Mistranslation		
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	
4.a	Palabras sin carga semántica							
4.a.1	Errores de carácter preposicional							
4.a.1.1	Preposiciones	1	(0,09)	12	(1,13)	5	(0,47)	18
4.a.1.2	Estructuras preposicionales					2	(0,19)	2
4.a.2	Determinantes	4	(0,38)	2	(0,19)			6
4.a.3	Errores de carácter adverbial							
4.a.3.1	Adverbios					3	(0,28)	3
4.a.3.2	Estructuras adverbiales					2	(0,19)	2
4.a.4	Conjunciones y partículas							

Tabla 222: Número y porcentaje total de errores en la cuarta categoría de palabras sin carga semántica en el motor RBMT2

9. Incidencia de cada categoría de errores en el corpus de análisis

Se marcaron 20 ediciones de preposiciones y 12 de estructuras preposicionales. En segundo lugar, se marcaron 6 errores en la traducción de determinantes y en tercer lugar 5 errores en la traducción de adverbios y estructuras adverbiales. En la ilustración a continuación se puede ver la distribución de los errores.

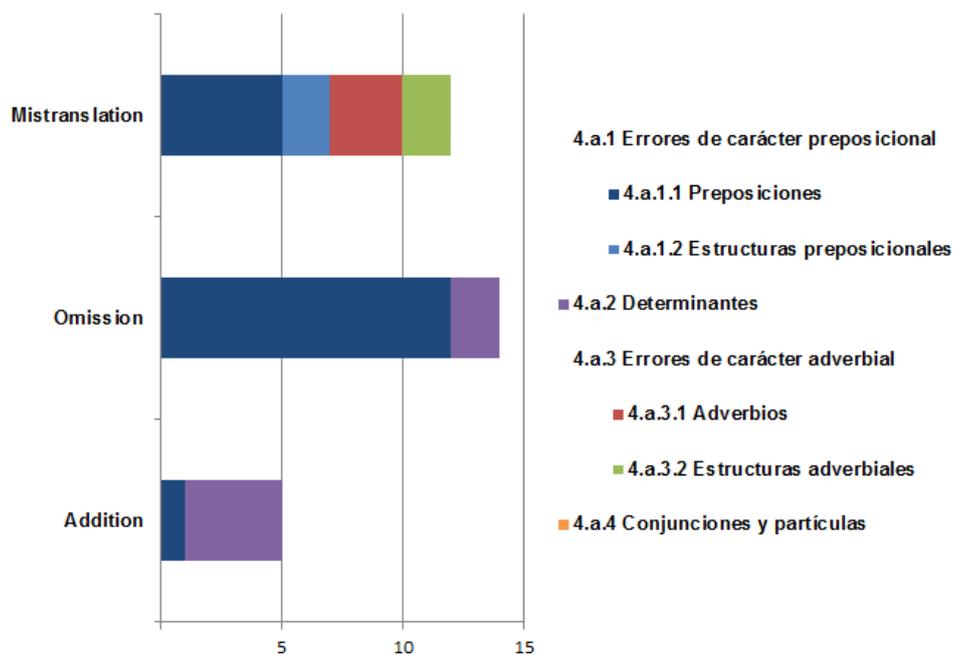


Ilustración 107: Relación del motor RBMT2 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica

La categoría en que se realizó un mayor número de ediciones en esta macrocategoría en general fue la que utilizamos para marcar los errores en las preposiciones o estructuras preposicionales (182 en total); en segundo lugar se realizaron más ediciones de determinantes (148 casos) y por último por errores que tienen que ver con los adverbios y estructuras adverbiales (40 casos); sin embargo la distribución en los motores del mismo tipo fue ligeramente diferente y no tan diferente como en la categoría anterior:

Motor	4.a.1 Preposiciones	4.a.2 Determinantes	4.a.3 Adverbios
SMT1	39	31	9
SMT2	23	34	2
RBMT1	107	69	24
RBMT2	20	6	5

Tabla 223: Categorías con más número de ediciones por motor

9. Incidencia de cada categoría de errores en el corpus de análisis

En la tabla anterior vemos que solo en un caso la categoría con más errores cambia en los motores del mismo tipo. Y esto se repite con las categorías en segundo y tercer lugar. Veamos ahora en la cuarta macrocategoría de palabras con carga semántica este fenómeno se repite.

9.3.4.2 Palabras con carga semántica

De la misma manera que en los apartados anteriores empezaremos con la ilustración en la que vemos la distribución total de errores en esta subcategoría, teniendo en cuenta todos los motores analizados. Aquí podemos ver que para casi todos los motores se realizaron la mayor cantidad de ediciones en la corrección de la traducción de sustantivos (4.b.1), verbos (4.b.3) y en editar errores de transferencia léxica de compuestos nominales (4.b.5).

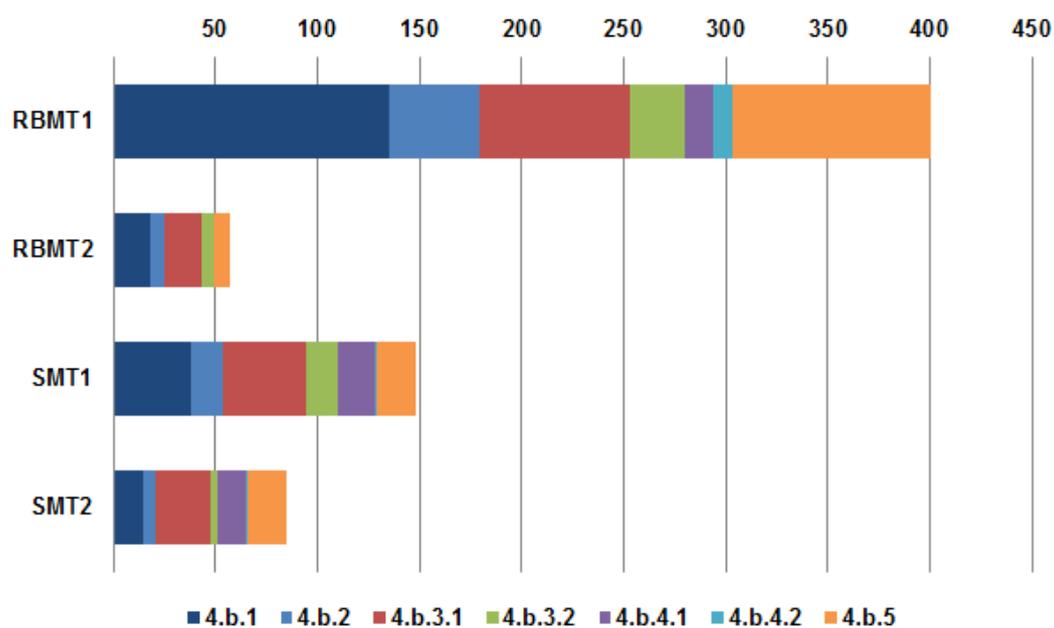


Ilustración 108: Relación tipo de motor categorías de la cuarta macrocategoría con carga semántica

Ahora veamos cómo es esta relación en cada motor. Empezaremos como en el apartado anterior con el motor SMT1. En primer lugar la tabla detallada de conteo de errores:

9. Incidencia de cada categoría de errores en el corpus de análisis

smt1		Addition		Omission		Mistranslation		Untranslated		Total Categoría
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
4.b.1	Sustantivos	4	(0,38)	2	(0,19)	28	(2,63)	4	(0,38)	38
4.b.2	Adjetivos	3	(0,28)	6	(0,56)	3	(0,28)	4	(0,38)	16
4.b.3	Verbos									
4.b.3.1	Verbos simples	4	(0,38)	22	(2,07)	13	(1,22)	1	(0,09)	40
4.b.3.2	Verbos compuestos			7	(0,66)	9	(0,85)			16
4.b.4	Errores relativos a compuestos									
4.b.4.1	Compuestos nominales					15	(1,41)	3	(0,28)	18
4.b.4.2	Fluidez del sintagma					1	(0,09)			1
4.b.5	Transferencia léxica					9	(0,85)	10	(0,94)	19

Tabla 224: Número y porcentaje total de errores en la cuarta macrocategoría (con carga semántica) en el motor SMT1

Aquí podemos observar que el mayor número de ediciones se dan en relación con la traducción de verbos (56 casos) y en la traducción de sustantivos (38 casos). Encontramos 19 errores relativos a compuestos y también 19 que tienen que ver con la transferencia léxica de compuestos nominales. Tenemos 78 casos de *mistranslation*, 37 casos de omisión, 22 casos de palabras que se dejaron sin traducir y 11 casos de adición; los otros valores no fueron relevantes para esta categoría. A diferencia de la macrocategoría anterior, se dan aquí muchos más casos de palabras que se dejaron sin traducir o *untranslated*.

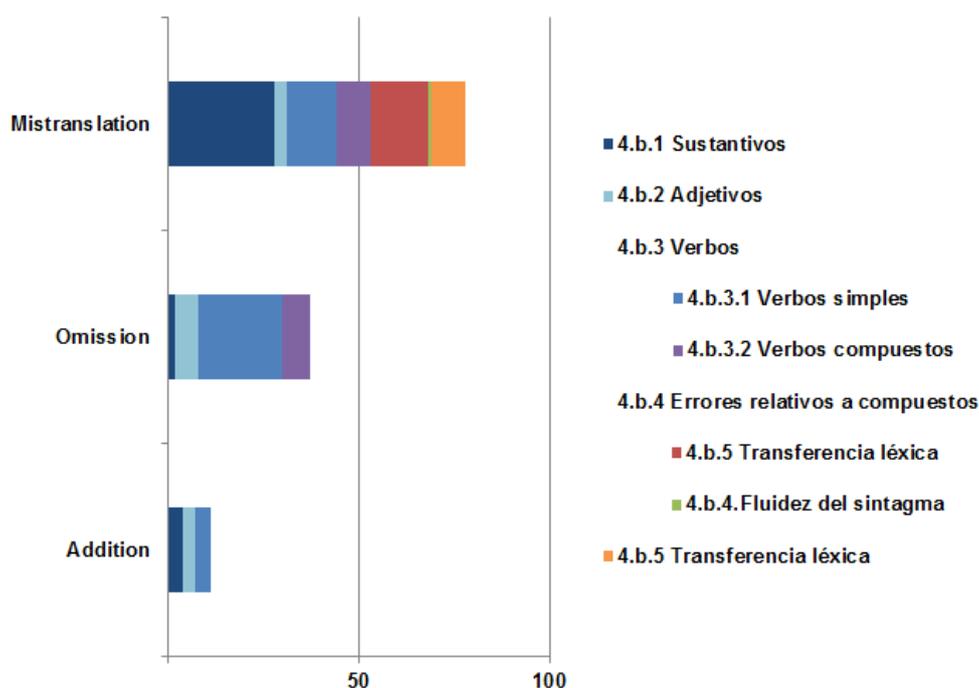


Ilustración 109: Relación del motor SMT1 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica

9. Incidencia de cada categoría de errores en el corpus de análisis

Ahora veamos cómo es la relación de ediciones marcadas en el SMT2. Podemos observar que el mayor número de ediciones también se dan en relación con la traducción de verbos (30 casos). Sin embargo en la traducción de sustantivos se presentan menos casos que en las propuestas del SMT1 (15 errores). También se marcaron 15 errores relativos a compuestos y 19 errores de transferencia léxica de compuestos nominales. El menor número de errores se presentó en la traducción de adjetivos. Tenemos 55 casos de *mistranslation*, 14 casos de omisión, 9 casos de adición y 7 casos de palabras que no se tradujeron; los otros valores no fueron tampoco con este motor relevantes para esta categoría.

smt2		Valor del Error								Total Categoría
		Addition		Omission		Mistranslation		Untranslated		
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
4.b.1	Sustantivos	2	(0,19)	1	(0,09)	10	(0,94)	2	(0,19)	15
4.b.2	Adjetivos	1	(0,09)			5	(0,47)			6
4.b.3	Verbos									
4.b.3.1	Verbos simples	6	(0,56)	9	(0,85)	11	(1,03)	1	(0,09)	27
4.b.3.2	Verbos compuestos			2	(0,19)	1	(0,09)			3
4.b.4	Errores relativos a compuestos									
4.b.4.1	Compuestos nominales					13	(1,22)	1	(0,09)	14
4.b.4.2	Fluidez del sintagma					1	(0,09)			1
4.b.5	Transferencia léxica			2	(0,19)	14	(1,32)	3	(0,28)	19

Tabla 225: Número y porcentaje total de errores en la cuarta categoría de palabras con carga semántica en el motor SMT2

Como en las propuestas del motor SMT1, en las propuestas seleccionadas del motor SMT2 la categoría en la que se marcaron más ediciones que ver con la traducción de verbos; en la siguiente ilustración podemos ver la distribución de todas las anotaciones realizadas:

9. Incidencia de cada categoría de errores en el corpus de análisis

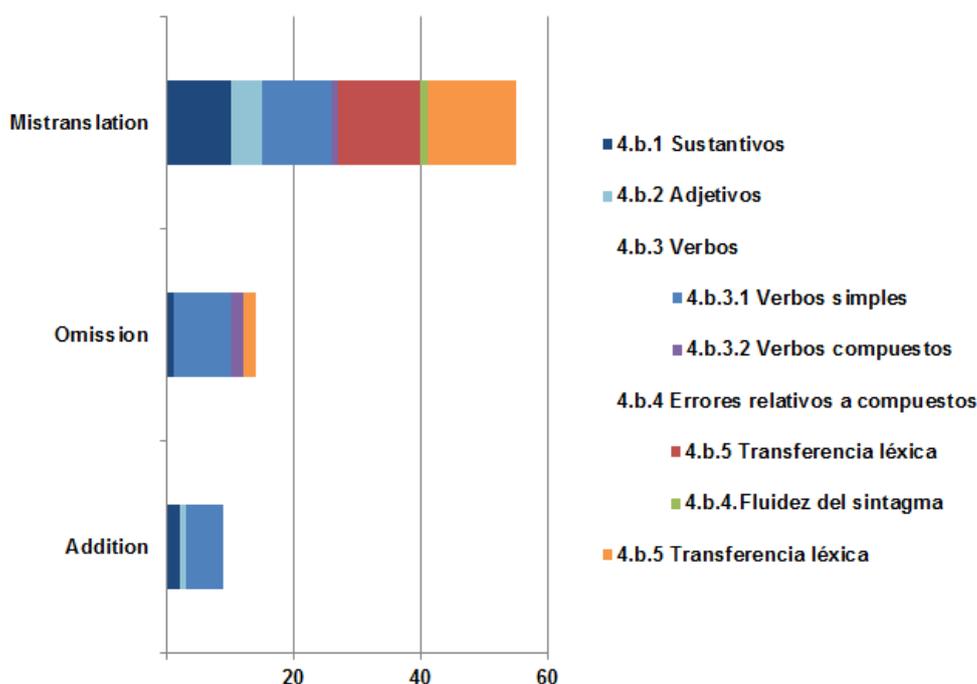


Ilustración 110: Relación del motor SMT2 en cada categoría de la cuarta macrocategoría en las palabras con carga semántica

Veamos si en los motores basados en reglas se presenta la misma relación:

Indicador	Categoría	Valor del Error								Total Categoría
		Addition		Omission		Mistranslation		Untranslated		
		Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
4.b.1 Sustantivos		3	(0,28)	4	(0,38)	103	(9,69)	25	(2,35)	135
4.b.2 Adjetivos						41	(3,86)	3	(0,28)	44
4.b.3 Verbos										
4.b.3.1 Verbos simples		5	(0,47)	1	(0,09)	60	(5,64)	7	(0,66)	74
4.b.3.2 Verbos compuestos		1	(0,09)	1	(0,09)	25	(2,35)			27
4.b.4 Errores relativos a compuestos										
4.b.4.1 Compuestos nominales						14	(1,32)			14
4.b.4.2 Fluidez del sintagma						9	(0,85)			9
4.b.5 Transferencia léxica				1	(0,09)	88	(8,28)	8	(0,75)	97

Tabla 226: Número y porcentaje total de errores en la cuarta categoría de palabras con carga semántica en el motor SMT2

En el motor RBMT1 se editaron 340 *mistranslation*, 6 omisiones y 9 adiciones. También encontramos un caso en que se editó una preposición que no se tradujo. Aquí se encontraron 35 casos en que se dejaron palabras si traducir. Encontramos una diferencia en relación con los motores estadísticos; el mayor número de ediciones realizadas se dio en la traducción de sustantivos y no —como en los motores estadísticos— en las traducciones de verbos. Se marcaron 101 ediciones en las traducciones de verbos y 23 errores relacionados con compues-

9. Incidencia de cada categoría de errores en el corpus de análisis

tos. Se realizaron 97 ediciones en las traducciones de compuestos nominales y 44 en las traducciones de adjetivos. En la ilustración a continuación se puede ver la distribución de los errores:

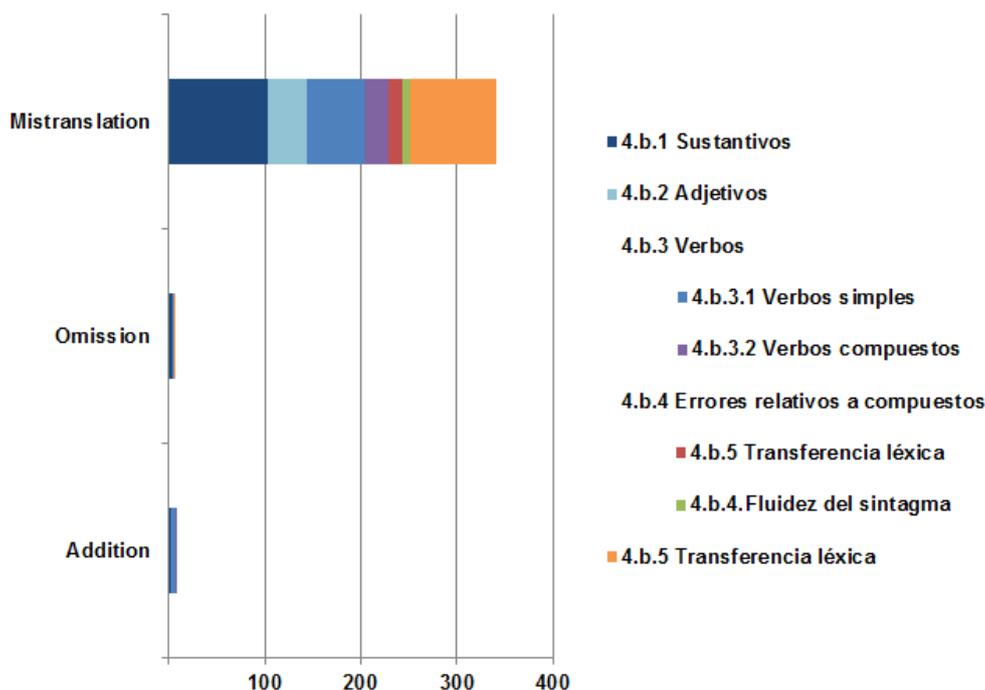


Ilustración 111: Relación del motor RBMT1 en cada categoría de la cuarta macrocategoría en palabras sin carga semántica

Las ediciones en el RBMT2 son del siguiente orden, en 40 casos se corrigieron errores por *mistranslation*, 4 omisiones, 2 adiciones y 8 casos de palabras que se dejaron sin traducir.

rbmt2	Indicador	Categoría	Valor del Error								Total Categoría
			Addition		Omission		Mistranslation		Untranslated		
			Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)	
	4.b.1	Sustantivos	1	(0,09)	1	(0,09)	10	(0,94)	6	(0,56)	18
	4.b.2	Adjetivos			2	(0,19)	4	(0,38)	1	(0,09)	7
	4.b.3	Verbos									
	4.b.3.1	Verbos simples	1	(0,09)	1	(0,09)	15	(1,41)	1	(0,09)	18
	4.b.3.2	Verbos compuestos					6	(0,56)			6
	4.b.4	Errores relativos a compuestos									
	4.b.4.1	Compuestos nominales									
	4.b.4.2	Fluidez del sintagma									
	4.b.5	Transferencia léxica					5	(0,47)	3	(0,28)	8

Tabla 227: Número y porcentaje total de errores en la cuarta categoría de palabras con carga semántica en el motor RBMT2

9. Incidencia de cada categoría de errores en el corpus de análisis

Al igual que en los motores estadísticos, la mayoría de las ediciones fueron correcciones de verbos (24 casos), luego se realizaron un mayor número de ediciones en sustantivos (18 casos) y un número similar de ediciones por transferencia léxica de compuestos nominales (8 casos) y por errores en la traducción de adjetivos (7 casos). A diferencia de los motores anteriores no se editaron en las propuestas de este motor errores relativos a compuestos.

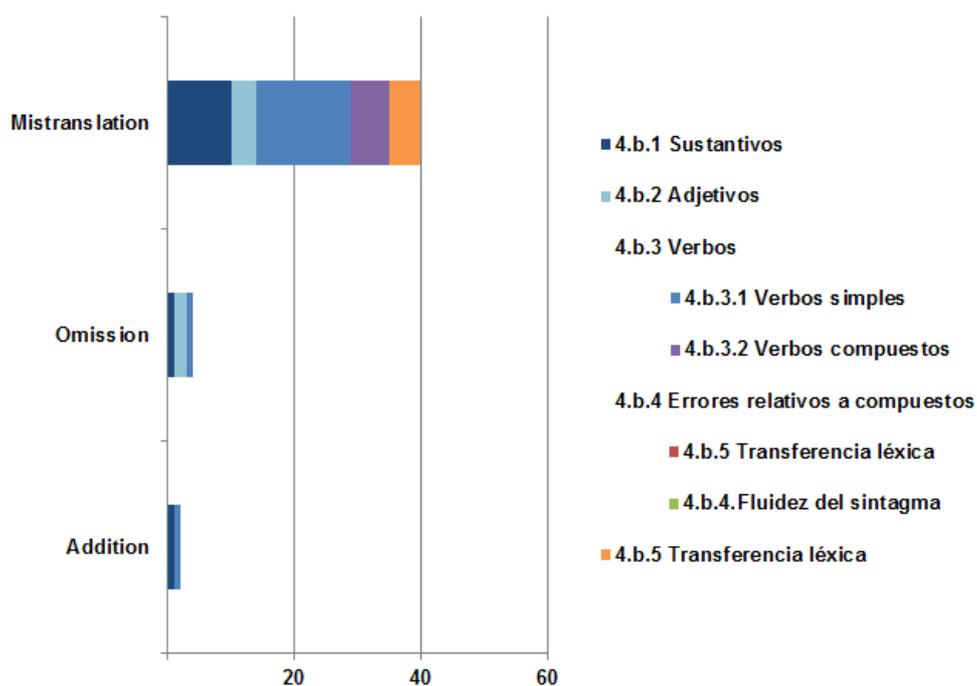


Ilustración 112: Relación del motor RBMT2 en cada categoría de la cuarta macrocategoría en palabras sin carga semántica

La categoría en que se realizó un mayor número de ediciones en esta macrocategoría en general fue la que utilizamos para marcar los errores en la traducción de verbos (211 en total); en segundo lugar se realizaron más ediciones en los sustantivos (206 casos) y por último por errores que tienen que ver con la transferencia léxica de compuestos nominales (143 casos); sin embargo la distribución en los motores del mismo tipo fue ligeramente diferente y en el motor en el que se seleccionaron más propuestas, se editaron más sustantivos que verbos:

9. Incidencia de cada categoría de errores en el corpus de análisis

Motor	4.b.1 Sustantivos	4.b.2 Adjetivos	4.b.3 Verbos	4.b.4 Compuestos	4.b.5 Transferencia léxica
SMT1	38	10	56	18	19
SMT2	15	6	30	15	19
RBMT1	135	44	101	23	97
RBMT2	18	7	24		8

Tabla 228: Categorías con más número de ediciones por motor

En la tabla anterior vemos que solo en un caso la categoría con más errores cambia en los motores del mismo tipo. Pero eso no se repite con las categorías en segundo y tercer lugar. Veamos ahora si en las categorías de la quinta macrocategoría también hay una categoría en la que se hayan realizado más ediciones en todos los motores.

9.3.5 Quinta macrocategoría errores relacionados con género del texto o la redacción del texto original

Esta macrocategoría agrupa los errores que se producen por problemas en la redacción o el diseño del original. En la ilustración a continuación podemos ver la distribución de las ediciones marcadas en todos los motores.

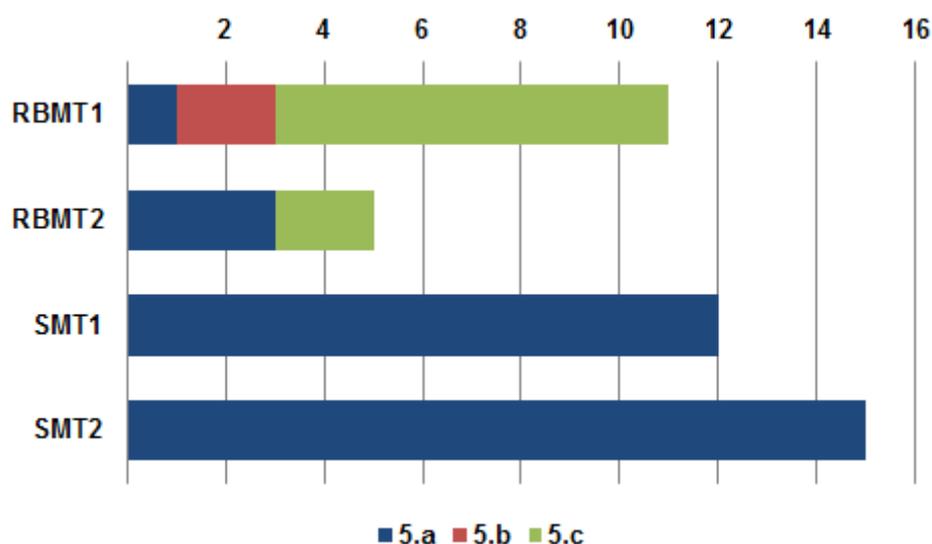


Ilustración 113: Relación tipo de motor categorías de la quinta macrocategoría

9. Incidencia de cada categoría de errores en el corpus de análisis

Ahora veremos cómo se distribuyen estos errores en cada motor en particular. En primer lugar, en el motor SMT1. Aquí fueron marcados 2 errores ortotipográficos y 12 errores por tratamiento incorrecto de las etiquetas:

smt1		Valor del Error					Total Categoría		
		Addition		Omission		Mistranlation		Untranslated	Questionable
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
5	Errores relacionados con el texto original								
5.a	Etiquetas							12	(25,00)
5.b	Estilo								
5.c	Ortotipográfico							2	(4,17)

Tabla 229: Número y porcentaje total de errores en la quinta categoría en el motor SMT1

En el motor SMT1 encontramos 15 errores por tratamiento erróneo de etiquetas y 2 errores ortotipográficos.

smt2		Valor del Error					Total Categoría		
		Addition		Omission		Mistranlation		Untranslated	Questionable
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
5	Errores relacionados con el texto original								
5.a	Etiquetas							15	(31,25)
5.b	Estilo								
5.c	Ortotipográfico								

Tabla 230: Número y porcentaje total de errores en la quinta categoría en el motor SMT2

En el motor SMT2 solo se editaron errores en las etiquetas.

rbmt1		Valor del Error					Total Categoría		
		Addition		Omission		Mistranlation		Untranslated	Questionable
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
5	Errores relacionados con el texto original								
5.a	Etiquetas			1	(2,08)				
5.b	Estilo								
5.c	Ortotipográfico	2	(4,17)	6	(12,50)			3	(6,25)

Tabla 231: Número y porcentaje total de errores en la quinta categoría en el motor RBMT1

Por el contrario, en el motor RBMT1 solo se presentó un problema relacionado con las etiquetas. Aquí se editaron tuvieron que editar errores que se dieron por problemas de orden ortotipográfico en el original.

rbmt2		Valor del Error					Total Categoría		
		Addition		Omission		Mistranlation		Untranslated	Questionable
Indicador	Categoría	Núm.	(%)	Núm.	(%)	Núm.	(%)	Núm.	(%)
5	Errores relacionados con el texto original								
5.a	Etiquetas							1	(2,08)
5.b	Estilo								
5.c	Ortotipográfico			2	(4,17)				

Tabla 232: Número y porcentaje total de errores en la quinta categoría en el motor RBMT2

9. Incidencia de cada categoría de errores en el corpus de análisis

En el caso del motor RBMT2 solo se presentaron 2 ediciones por problemas relacionado con las etiquetas. Aquí se editaron solo 2 errores que se dieron por problemas de orden ortotipográfico en el original.

La categoría en que se realizó un mayor número de ediciones en esta macrocategoría en general fue la que utilizamos para marcar el tratamiento erróneo de las etiquetas (31 errores); en segundo lugar se realizaron más ediciones por problemas ortotipográficos en el original (15) y en último lugar se editaron solo 2 casos por problemas de estilo.

Motor	5.a Etiquetas	5.b Estilo	5.c Ortotipografía
SMT1	12		2
SMT2	15		
RBMT1	1	2	11
RBMT2	3		2

Tabla 233: Categorías con más número de ediciones por motor

En esta categoría la distribución de las ediciones es muy similar, solo el motor RBMT1 presenta un orden diferente. En este motor no se tuvieron que editar errores de tratamiento erróneo de las etiquetas, sino problemas de orden ortotipográfico.

Para determinar por qué los poseedores se deciden por la propuesta de un motor o por la propuesta de otro compararemos en el siguiente aparte en detalle el proceso de selección teniendo en cuenta dos categorías primordiales, es decir, dos de las categorías en las que se realizaron más ediciones, nos referimos a la categoría 4.b.1 en la que clasificamos las ediciones de los sustantivos y la categoría 3.b en la que clasificamos los problemas de flexión verbal.

9.4 Incidencia por motor y valores

Como hemos visto en los resultados anteriores el mayor número de ediciones se realizaron en la traducción de verbos (211 casos), sustantivos (206), preposiciones y construcciones preposicionales (182) y en la transferencia léxica de compuestos nominales (143). Por ello queremos ver en detalle qué valores se asignaron a las ediciones en cada uno de estos casos. Empezaremos de acuerdo al número de ediciones.

9. Incidencia de cada categoría de errores en el corpus de análisis

9.4.1 Valores asignados a las ediciones de verbos simples y compuestos

Para hacernos una idea clara acerca de qué tipo de errores prefieren corregir los poseedores tendremos en cuenta los valores asignados en cada motor. Como vemos a continuación la mayoría de las ediciones realizadas cuando se encontraron errores en los verbos se debe a la traducción errónea de estos. En segundo lugar se editaron las omisiones de su traducción y en tercer lugar se editaron los casos en que se añadieron verbos que no estaban en el original. En la tabla a continuación podemos ver el número de ediciones teniendo en cuenta el valor y el tipo de motor, entre paréntesis se encuentra el porcentaje con respecto al total de ediciones realizadas en total en la macrocategoría.

	Addition	Omission	Mistranslated	Untranslated
SMT1	4 (0,38 %)	29 (2,73 %)	22 (2,07 %)	1 (0,09 %)
SMT2	6 (0,56 %)	11 (1,03 %)	12 (1,13 %)	1 (0,09 %)
RBMT1	6 (0,56 %)	2 (0,19 %)	85 (8,00 %)	7 (0,66 %)
RBMT2	1 (0,09 %)	1 (0,09 %)	21 (1,98 %)	1 (0,09 %)
Total	17 (8,09 %)	43 (20,47 %)	140 (66,66 %)	10 (4,76 %)

Tabla 234: Número y porcentaje de los valores asignados a las ediciones de los verbos

En el próximo apartado haremos en mismo listado con las ediciones de sustantivos, para ver si se repite la distribución de los valores asignados a las ediciones de los verbos.

Con respecto a los verbos hemos realizado una comparación adicional. En 122 de los segmentos se editaron verbos simples y/o compuestos; ya que en el corpus original teníamos las propuestas de traducción de todos los motores pudimos comparar en estos casos las propuestas de los motores descartados en la posesición. Así vimos a) si se dan los mismos errores o no en las traducciones de los segmentos con los otros motores, b) si los segmentos de los otros motores contienen más problemas que el segmento elegido. Una vez realizada esta comparación encontramos efectivamente algunos casos en que en los segmentos elegidos se daba el mismo error, pero presentaban menos errores de otro tipo. En la mayoría de los casos vimos que se reproducía el error o se presentaban más errores. Aquí un ejemplo, en el segmento del RBMT1 que se selecciona para la posesición no se traduce correctamente el verbo *ändern* (se traduce como cambian, en la posesición se prefirió modificar). En el segmento del otro motor

9. Incidencia de cada categoría de errores en el corpus de análisis

se omite la traducción del verbo. Encontramos otros casos similares en que el poseedor prefirió corregir la traducción y la conjugación del verbo.

	Motor	LV	
Original			<i>Standardmäßig ist die Registerkarte Anwender ausgewählt, hier können Sie einen neuen Anwender erstellen oder die Eigenschaften des Benutzerprofils ändern, mit dem Sie einen Zugang zum System erhalten haben.</i>
Propuesta seleccionada	RBMT1	87	Le es estandarizada la tarjeta de registro a usuario seleccionado, aquí pueden elaborar Usted un usuario nuevo o las propiedades del perfil de usuarios con el cual ha recibido una entrada al sistema cambian .
Propuesta con los mismos o menos errores	SMT1	120	Es el registro de los usuarios de la tarjeta normalizada seleccionado a elaborar un nuevo usuario o las características del usuario, con el perfil de un acceso al sistema.
Posedición			La página de registro está seleccionada de manera estándar, aquí puede crear usted un usuario nuevo o modificar las propiedades del perfil de usuario con el cual ha recibido un acceso al sistema.

Tabla 235: Ejemplo de edición de errores en la traducción de verbos

En la propuesta del motor SMT1, que no fue seleccionada para la posedición no solo se tienen que corregir otros errores sino que además se omitió la traducción del verbo. Teniendo en cuenta este ejemplo y la tabla en la que ilustramos los valores asignados a las ediciones realizadas podemos ver que los poseedores prefieren corregir traducciones erróneas a traducir de cero.

Ahora veremos si lo mismo se da en las ediciones de sustantivos, compuestos nominales, verbos simples y compuestos y verbos separables.

9.4.2 Valores asignados a las ediciones de sustantivos

También en el caso de los sustantivos la mayoría de las ediciones se dieron por la necesidad de corregir una traducción incorrecta. No podemos corroborar que los motores cometan más errores de esta categoría pero sí que el poseedor prioriza segmentos con este problema en concreto por delante de otros errores. A diferencia de los valores que encontramos en las ediciones de los verbos, los valores *addition* u *omission* se presentan en menos casos, pero

9. Incidencia de cada categoría de errores en el corpus de análisis

se dan más casos de ediciones de *untranslated*. En la tabla a continuación vemos el número de ediciones en las que seleccionamos este valor y entre paréntesis el porcentaje que este número tiene en relación con el total de errores marcados en la traducción de sustantivos.

	Addition	Omission	Mistranslated	Untranslated
SMT1	4 (0,38 %)	2 (0,19 %)	28 (2,63 %)	4 (0,38 %)
SMT2	2 (0,19 %)	1 (0,09 %)	10 (0,94 %)	2 (0,19 %)
RBMT1	3 (0,28 %)	4 (0,38 %)	103 (9,69 %)	25 (2,35 %)
RBMT2	1 (0,09 %)	1 (0,09 %)	10 (0,94 %)	6 (0,56 %)
Total	10 (4, 85%)	8 (3, 88 %)	151 (73, 30 %)	37 (17, 96 %)

Tabla 236: Valores de los errores en la traducción de sustantivos

Con respecto a los sustantivos también hemos realizado la comparación que hicimos con los verbos. En 173 de los segmentos se editaron sustantivos; ya que en el corpus original teníamos las propuestas de traducción de todos los motores pudimos comparar en estos casos las propuestas de los motores descartados en la posesición. Así vimos a) si se dan los mismos errores o no en las traducciones de los segmentos con los otros motores, b) si los segmentos de los otros motores contienen más problemas que el segmento elegido. Una vez realizada esta comparación encontramos efectivamente algunos casos en que las propuestas se encontraba el mismo error, pero se presentaban menos errores de otro tipo. Por ello miraremos detalladamente estos casos:

	Motor	LV	
Original			<i>Manuelle Deinstallation</i>
Propuesta seleccionada	RBMT1	4	Deinstallation manual
Propuesta con los mismos o menos errores	SMT1	4	Deinstallation manual
Posedición			Desinstalación manual
Original			<i>Automatische Deinstallation des Produkts</i>

9. Incidencia de cada categoría de errores en el corpus de análisis

Propuesta seleccionada	SMT1	4	Deinstallation automática del producto ¹²
Propuesta con los mismos o menos errores	SMT2	4	Deinstallation automática del producto
Posedición			Desinstalación automática del producto
Original			<i>In diesem Fall können Sie nach dem Klicken auf das unbekannte Gerät das jeweilige Modell auswählen.</i>
Propuesta seleccionada	SMT1	45	En este caso, usted puede tras el clic en el aparato desconocidos escoger el modelo de cada país.
Propuesta con los mismos o menos errores	SMT2	43	En este caso, puede usted tras el clic sobre el aparato desconocido de elegir el modelo.
Posedición			En este caso, después de hacer clic en el equipo desconocido, usted puede escoger el modelo correspondiente.
Original			<i>Möglicherweise wird im Rahmen des Setupvorgangs die folgende Nachricht angezeigt.</i>
Propuesta seleccionada	RBMT2	58	Aquellos posiblemente al amparo de es señalado del suceso de setup noticia siguiente.
Propuesta con los mismos o menos errores	RBMT1	55	Posiblemente se indica en el marco del Setupvorgangs la noticia siguiente.
Posedición			Es posible que durante el proceso de instalación se muestre el siguiente mensaje:
Original			<i>In der Registerkarte Konfiguration gibt es Informationen über die allgemeinen- und Postserver:</i>
Propuesta seleccionada	SMT1	39	En la tarjeta de registro configuración hay información sobre la allgemeinen- y servidores postales:
Propuesta con los mismos o menos errores	RBMT1	36	En la tarjeta de registro configuración hay informaciones sobre éstos general y correo servidores:
Posedición			En la ficha de registro "configuración" hay información sobre los ser-

¹² Aquí las propuestas de todos los motores eran idénticas.

9. Incidencia de cada categoría de errores en el corpus de análisis

			vidores generales y de correo:
Original			<i>Registerkarte LOGs</i>
Propuesta seleccionada	SMT1	6	Tarjeta de registro LOGs
Propuesta con los mismos o menos errores	RBMT1	6	Tarjeta de registro LOGs
Posedición			Ficha de registro LOGs
Original			<i>Von jedem Punkt des Programms aus können Sie stets durch einen Klick auf jede einzelne der Optionen des Hauptmenüs auf der linken Seite zugreifen, daher klicken Sie nun auf Verwaltung und danach auf Geräte.</i>
Propuesta seleccionada	SMT1	68	Desde cualquier punto del programa puede usted siempre por un clic en cada una de las opciones del Hauptmenüs acceden a la izquierda, por lo tanto, Internet y ahora de la administración y los aparatos.
Propuesta con los mismos o menos errores	RBMT1	59	Desde cada punto del programa puede acceder siempre a cada una de las opciones del menú principal por un clic en el lado izquierdo, por eso hace clic ahora sobre aparatos en administración y a eso.
Posedición			Desde cualquier punto del programa puede siempre acceder con un clic a cada una de las opciones del menú principal situado a la izquierda, por lo tanto, haga clic ahora en "administración" y después en "equipos".

Tabla 237: Propuestas similares rechazadas en comparación con las seleccionadas para la posedición

Aunque no sabemos de qué manera les aparecían las opciones en pantalla a los poseedores los textos, parece que se presentaron en primer lugar las propuestas del SMT1. Esto, si bien hace que no podamos encontrar criterios que nos ayuden a ver qué tipos de errores prefiere corregir el poseedor, relativiza el hecho de que se hayan poseditado más segmentos procedentes del motor RBMT1 debido a que siempre se hubiese presentado como primera alternativa.

Vemos si la distancia de teclado nos revela otro tipo de información. Con ayuda de la distancia de teclado detectamos 26 casos en que, a pesar de que la distancia Levenshtein era menor en otra propuesta, se escogió una en la que se hicieron finalmente más ediciones.

9. Incidencia de cada categoría de errores en el corpus de análisis

En el ejemplo a continuación podemos ver que los poseedores prefieren, por ejemplo, corregir traducciones erróneas pero no corregir las etiquetas:

	Motor	LV	
Original			"<field name="\$paratext"/>" <i>auf Seite</i> <field name="\$pagenum"/>
Propuesta seleccionada	RBMT1	8	"<field name="\$paratext"/>" en lado <field name="\$pagenum"/>
Propuesta de otro motor	RBMT2	6	""<field name="\$paratext"/>" en lado <field name="\$pagenum"/>
Propuesta de otro motor	SMT1	46	" las cosas field name = " \$texto " / me" en la página de hacer las cosas field name = "\$pagenum" / me
PE			"<field name="\$paratext"/>" en la página <field name="\$pagenum"/>;

Tabla 238: En la PE se evita la edición de etiquetas

En la tabla anterior vemos que la propuesta del motor RBMT2 tiene un valor LV menor que la de la propuesta seleccionada, pero aunque se comete el mismo error, el poseedor selecciona la propuesta del motor RBMT1 quizás porque en esta propuesta no se dañan las etiquetas. En la propuesta del SMT1 la traducción es correcta, pero el poseedor parece que evita el arreglar las etiquetas.

Otro tipo de corrección que se prefirió fue el editar una palabra mal traducida a traducir una *untranslated*, incluso en un caso como el siguiente, donde la corrección de la traducción hubiese implicado menos esfuerzo:

	Motor	LV	
Original			<i>Unter Daten finden Sie die von der Serversoftware erfassten Daten der verschiedenen Gerätesonden.</i>
Propuesta seleccionada:	SMT1	79	Bajo los datos encontrarán ustedes los datos recopilados mercancía con mucha frecuencia por el servidor de los diferentes Gerätesonden .
Propuesta de otro motor:		52	Entre datos encuentra los datos comprendidos por el software de servidor de las sondas de aparato diferentes.
PE:			En "Datos" encontrará los datos recopilados por el software del servidor referentes a las distintas sondas del aparato .

Tabla 239: En la PE se evita la traducción de *untranslated*

9. Incidencia de cada categoría de errores en el corpus de análisis

Sin embargo esta apreciación no se puede generalizar. En el siguiente caso nos dimos cuenta de que no se traduce correctamente el sustantivo *Maustaste* (se traduce solo ‘ratón’, en la posesición se empleó ‘botón del ratón’). Al parecer el poseedor prefiere realizar pocos cambios en orden sintáctico a tener que editar el sustantivo, lo que es necesario en ambos casos, y además tener que traducir de cero los elementos no traducidos en la propuesta con un mejor valor según LV:

	Motor	LV	
Original			<i>Wenn Sie die Maustaste loslassen, wird die Grafik neu berechnet und der ausgewählte Bereich angezeigt.</i>
Propuesta seleccionada:		47	Si usted el ratón "en esta decisión, la gráficos de nuevo y el ámbito seleccionados.
Propuesta de otro motor:		36	Si suelta el comando del ratón , es neu berechnet el gráfico e indicado el área seleccionada.
PE:			Si suelta el botón del ratón , se vuelve a calcular el gráfico y se muestra el área seleccionada.

Tabla 240: En la PE prefiere realizar ediciones en el orden sintáctico

La preferencia por editar problemas en verbos y sustantivos en lugar de editar errores sintácticos, de etiquetas o de otro tipo podría tratarse de un fenómeno extrapolable a cualquier otra combinación lingüística. Veamos a continuación qué sucede en el caso de los compuestos nominales, una unidad léxica propia del alemán.

9.4.3 Compuestos nominales

Ya desde el marco teórico llamamos la atención sobre la problemática de la traducción de los compuestos nominales. En el conteo de errores nos dimos cuenta que este tipo de errores se tiene que editar frecuentemente y que generalmente se dejan sin traducir o se traducen erróneamente, porque el orden sintáctico no es correcto; no encontramos en ninguna de las propuestas seleccionadas para la posesición ningún caso en el que se añadieran o se omitieran compuestos nominales:

9. Incidencia de cada categoría de errores en el corpus de análisis

	Addition	Omission	Mistranslated	Untranslated
SMT1			16 (1,51 %)	3 (0,28 %)
SMT2			14 (1,32%)	1 (0,09 %)
RBMT1			23 (2,16 %)	
RBMT2				

Tabla 241: Valores de los errores sintácticos en la traducción de compuestos nominales

En la ilustración a continuación se ve claramente como el número de propuestas en las que se editaron las traducciones de compuestos nominales es muy similar. Y solo en un motor (RBMT2) no se seleccionaron propuestas que tuviesen este tipo de error. También vemos que no se marcaron casos de adición u omisión de palabras:

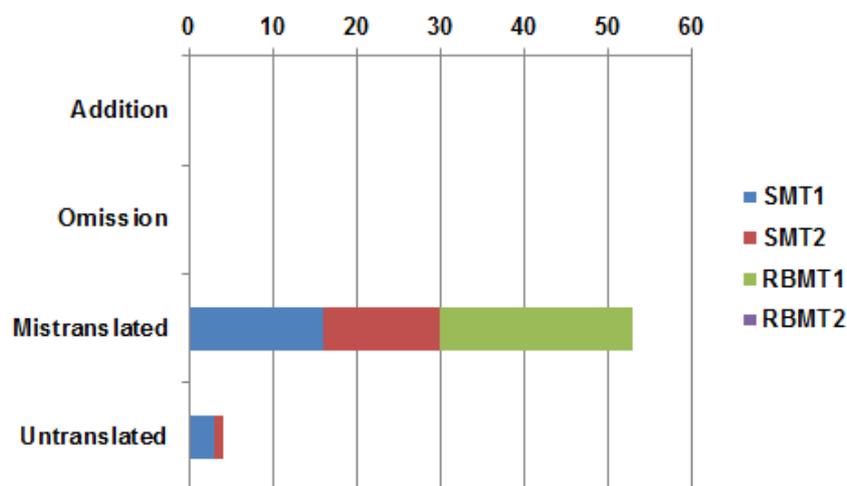


Ilustración 114: Relación de los valores de errores sintácticos en la traducción de compuestos nominales

No solo se presentaron problemas de orden sintáctico en la traducción de compuestos nominales, sino también problemas en la transferencia léxica. En la tabla a continuación vemos su distribución:

	Addition	Omission	Mistranslated	Untranslated
SMT1			9 (0,85 %)	10 (0,94 %)
SMT2		2 (0,19 %)	14 (1,32%)	3 (0,28 %)
RBMT1		1 (0,09)	88 (8,28)	8 (0,75)
RBMT2			5 (0,47)	3 (0,28)

Tabla 242: Valores de los errores en la transferencia léxica de compuestos nominales

9. Incidencia de cada categoría de errores en el corpus de análisis

Aquí si marcamos dos casos de omisión de traducción de los elementos del compuesto nominal. Pero la mayoría de los errores se catalogaron como *mistranslation*.

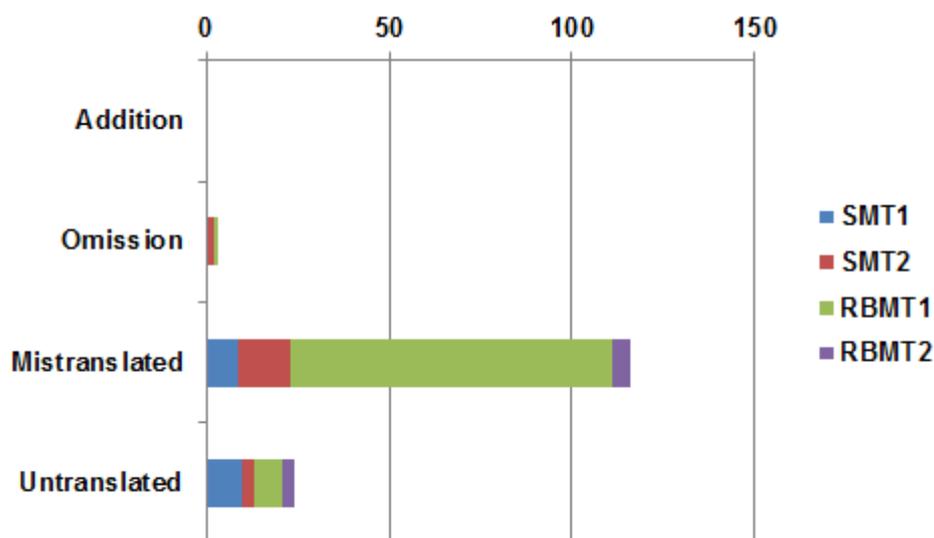


Ilustración 115: Relación de los valores de errores léxicos en la traducción de compuestos nominales

En las propuestas con errores de transferencia léxica en los compuestos nominales nos encontramos con un comportamiento diferente al que registramos en la traducción de sustantivos, aquí se prefirió corregir las etiquetas si la traducción del compuesto nominal era más adecuada:

	Motor	LV	
Original			<Frm id="17"/> Nachtabsenkung:
Propuesta seleccionada:	SMT2	41	Sucedá Frm ID = "17". / suceda. B: Reducción de la noche suceda / b.
Propuesta de otro motor:	RBMT1	16	<Frm id="17"/> Rebajamiento nocturno:
PE:			Frm id="17"/> Reducción nocturna:
Original			Der Wechsel auf Hocheffizienzpumpen amortisiert sich schneller als jede andere energetische Maßnahme.
Propuesta seleccionada:	SMT1	35	Las cosas Tab / me el cambio a la alta eficiencia se amortizado inyectar más rápido que cualquier otra medida energética.

9. Incidencia de cada categoría de errores en el corpus de análisis

	Motor	LV	
Propuesta de otro motor:	RBMT1	22	<Tab/>El cambio en bombas de viva eficiencia se amortiza rápido que cada otra medida energética.
PE:			<Tab/>El cambio a bombas de alta eficiencia se amortiza más rápido que cualquier otra medida energética.

Tabla 243: Tratamiento de las etiquetas y edición de compuestos nominales

En algunas ocasiones no hubo un criterio claro que nos dejara identificar por qué el poseedor se decidió por la propuesta de un motor, si otra —en la que se reproducía el mismo error— también se hubiera podido seleccionar:

	Motor	LV	
Original			<i>Bei einem Doppelklick auf den Gerätenamen erhalten Sie Zugang zu ihrem Parametersatz:</i>
Propuesta seleccionada:	SMT1	45	En un doble clic o a los aparatos nombre tendrá acceso a su parámetros frase :
Propuesta de otro motor:	RBMT1	45	En un clic doble en los nombres de aparato recibe entrada a su frase de parámetro :
PE:			Haciendo doble clic en el nombre de los aparatos tendrá acceso a su conjunto de parámetros :

Tabla 244: Propuestas con el mismo valor LV de dos motores diferentes

En algunos casos, como los de siguiente ejemplo podemos ver que había al menos otra propuesta en la que como en la seleccionada para la posesión no solo encontramos el error en la transferencia léxica del compuesto nominal sino además problemas de otro tipo. En algunos casos el valor LV era igual o incluso menor. Aquí, por ejemplo, en la propuesta seleccionada para la posesión encontramos en dos adjetivos errores de concordancia en la flexión de género y número, no se traduce correctamente la preposición 'bei' tampoco se traduce correctamente el verbo 'geht' y se deja sin traducir el anglicismo. La otra propuesta presenta errores de puntuación, tampoco se traduce correctamente la preposición, ni el verbo, pero no tiene los errores de concordancia de género. En la propuesta con un LV menos falta la traducción del verbo y se añade un sustantivo. Con este ejemplo queremos mostrar que los poseedores prefieren editar —en las propuestas con problemas en la transferencia léxica de

9. Incidencia de cada categoría de errores en el corpus de análisis

compuestos nominales— propuestas en las que no se añaden palabras que no se encuentran en el original o en las que no se omitían palabras, así estas no se hayan traducido.

	Motor	LV	
Original			<i>Bei sehr kleinen Volumenströmen geht die Pumpe hierzu in einen hydraulischen Standby.</i>
Propuesta seleccionada:	RBMT1	48	En corrientes de volumen muy pequeñas va la bomba a eso a una disponibilidad hidráulica.
Propuesta de otro motor:	SMT2	45	Cuando los volúmenes muy pequeños, la bomba en un televisor hidráulicos al respecto.
PE:			Con flujos volumétricos muy pequeños la bomba se pone en modo de espera hidráulico.

Tabla 245: En la PE se seleccionan propuestas sin adiciones u omisiones

Otro ejemplo interesante se da con la puntuación. A pesar de que en el corpus se encontraban propuestas en las que no se reproducía el error de transferencia léxica, los poseedores prefieren editar las propuestas que tuvieran la puntuación correcta:

	Motor	LV	
Original			<i>Bis hierher haben wir alle Optionen in Bezug auf die Systemverwaltung eingesehen, im Folgenden werden wir uns mit den Optionen zur Geräteverwaltung beschäftigen.</i>
Propuesta seleccionada:	SMT1	34	Hasta aquí hemos todas las opciones en relación con el sistema de administración aceptó, vamos a abordar las opciones para la gestión de los aparatos .
Propuesta de otro motor:	SMT2	33	Hasta aquí hemos todas las opciones en relación con el sistema de administración aceptó a continuación nos con las opciones para la gestión de equipos .
PE:			Hasta aquí hemos visto todas las opciones en relación con el sistema de administración, vamos a abordar a continuación las opciones para la gestión de los equipos .

Tabla 246: Puntuación correcta

9. Incidencia de cada categoría de errores en el corpus de análisis

Si bien en las propuestas en que las que se había traducido erróneamente un sustantivo se prefirió no editar las propuestas que tuviesen problemas con las etiquetas, vemos ahora que en lugar de editar la traducción errada de un compuesto nominal, los poseedores prefieren corregir el tratamiento erróneo de las etiquetas.

Una tendencia si se confirmó, los poseedores prefieren corregir traducciones incorrectas (*mistranslation*) a traducir palabras que se hayan omitido e incluso eliminar adiciones.

9.4.4 Verbos separables

Encontramos 17 casos en los que se editó la traducción de verbos separables. En la siguiente tabla vemos cómo se distribuyeron que valores le asignamos a los cambios realizados por los poseedores.

	Addition	Omission	Mistranslated	Untranslated
SMT1		2 (4,76 %)	1 (2,38 %)	
SMT2			2 (4,76 %)	
RBMT1			10 (23,81)	
RBMT2			1 (2,38 %)	

Tabla 247: Valores de los errores en la transferencia de verbos compuestos separables

En las propuestas que se seleccionaron para la posesición que tenían errores en la traducción de verbos separables no encontramos casos en los que asignamos los valores de *addition* o *untranslated*.

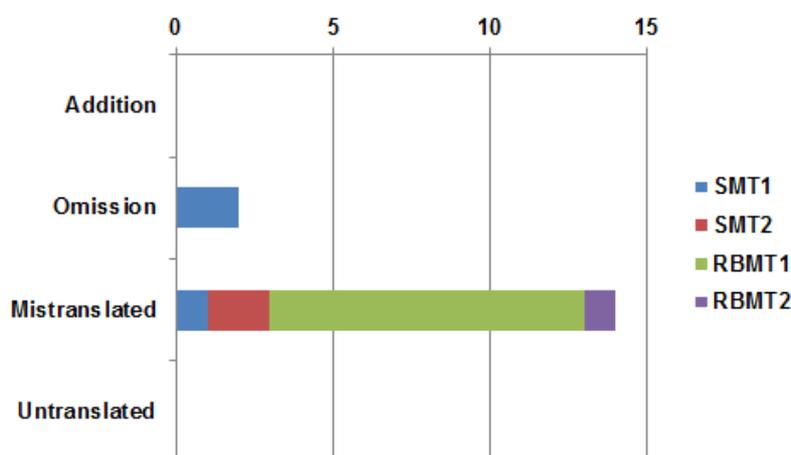


Ilustración 116: Relación de los valores la traducción de verbos alemanes separables

9. Incidencia de cada categoría de errores en el corpus de análisis

Como en las ediciones en las traducciones de sustantivos y compuestos nominales, también encontramos casos en los que el poseedor seleccionó la propuesta de un motor, a pesar de que otra propuesta tenía un valor LV menor. Veremos a continuación cada uno de los cuatro casos por separado:

	Motor	LV	
Original			<ut/>Legen Sie die bereitgestellte CD in das CD-ROM-Laufwerk Ihres Computers ein.
Propuesta seleccionada:	RBMT2	47	<ut/>Macerar EL CD en la unidad de CD-ROM de Su ordenador aprontado.
Propuesta de otro motor:	RBMT1	43	<ut/>Formule el CD preparado al mecanismo de ROMA DE CD de su ordenador.
PE:			<ut/>Introduzca el CD puesto a su disposición en la unidad de CD-ROM de su ordenador.

Tabla 248: Ejemplo de las ediciones en errores de verbos separables en dos propuestas diferentes

En el ejemplo anterior no se traduce correctamente el verbo compuesto separable *einlegen*, en la propuesta seleccionada para la posesición se traduce como ‘macerar’ la traducción correcta sería ‘introduzca’. La propuesta del RBMT1 también traduce el verbo de manera errónea, pero además traduce erróneamente la composición ‘CD-ROM-Laufwerk’ y añade el sustantivo ‘mecanismo’, que no se encuentra en el original. Esto corrobora lo que vimos en los compuestos nominales, los poseedores prefieren editar traducciones erradas a eliminar adiciones.

	Motor	LV	
Original			Beim ersten Programmzugang geben Sie das Wort <ut/>superuser<ut/> ein, sowohl für den <ut/>Login<ut/>, als auch für das Passwort, in Kleinbuchstaben und ohne Leerstellen.
Propuesta seleccionada:	SMT1	120	En el primer programa acceso la palabra suceda ut / usuarios me suceda ut / me, tanto para los suceda ut / me Login suceda ut / me, como también para el uso de la palabra, en la letra pequeña y sin vacíos.
Propuestas de otros motores:	RBMT1	56	En la primera entrada de programas da la palabra <ut/>superuser<ut/> un, tanto para ése <ut/>Login<ut/>, que también para la contraseña, en letras minúsculas y sin blancos.

9. Incidencia de cada categoría de errores en el corpus de análisis

	Motor	LV	
	SMT2	98	En el primer programa de acceso a la palabra suceda. / ut ut suceda. / Super usuarios, tanto para el suceda. / ut ut Login suceda. /, como para la contraseña, en la letra pequeña y sin vacíos.
PE:			En el primer acceso al programa introduzca la palabra <ut/>superuser<ut/>, tanto para <ut/>iniciar sesión<ut/> como para la contraseña, en letras minúsculas y sin espacios.

Tabla 249: Ejemplo de las ediciones en errores de verbos seapbles en dos propuestas diferentes

En el ejemplo anterior en la propuesta seleccionada falta la traducción del verbo separable *eingeben*. A pesar de que hay errores por tratamiento erróneo de las etiquetas el poseedor opta por esta opción. Aquí no encontramos una razón clara que explique esta decisión; la propuesta del SMT2 también tiene problemas por tratamiento erróneo de las etiquetas y en la propuesta del RBMT1 con el mejor valor LV no se traduce correctamente el verbo, pero como el valor lo indica hubiese sido más rápido corregirla.

El tercer caso no se traduce correctamente el verbo compuesto separable *ein(ge)sehen haben*, se traduce como ‘aceptó’, la opción correcta sería correcta sería ‘hemos visto’.

	Motor	LV	
Original			<i>Bis hierher haben wir alle Optionen in Bezug auf die Systemverwaltung eingesehen, im Folgenden werden wir uns mit den Optionen zur Geräteverwaltung beschäftigen.</i>
Propuesta seleccionada:	SMT1	34	Hasta aquí hemos todas las opciones en relación con el sistema de administración aceptó , vamos a abordar las opciones para la gestión de los aparatos.
Propuesta de otro motor:	SMT2	33	Hasta aquí hemos todas las opciones en relación con el sistema de administración aceptó a continuación nos con las opciones para la gestión de equipos.
PE:			Hasta aquí hemos visto todas las opciones en relación con el sistema de administración, vamos a abordar a continuación las opciones para la gestión de los equipos.

Tabla 250: Ejemplo de propuestas similares, con puntuación diferente

9. Incidencia de cada categoría de errores en el corpus de análisis

El error se da en ambas propuestas, pero el poseedor se decide por la propuesta en que no tienen que añadir la coma. Es decir, aquí también vemos que los poseedores prefieren no tener que hacer ajustes en la puntuación:

Veamos ahora el último caso. Aquí falta la traducción del verbo compuesto separable *zugrundeliegen*, que también debe ser ‘se basan’. En ambas propuestas falta la traducción del verbo, sin embargo se escoge la del SMT1 y a pesar del error de puntuación. La única diferencia notable es que en la propuesta con un mejor valor LV la traducción del sustantivo *Strompreis* está en plural ‘precios’ y no ‘precio’ que es la traducción correcta.

	Motor	LV	
Original			<i>Den Berechnungen liegt ein mittlerer Strompreis von 19 Cent/kWh zugrunde.</i>
Propuesta seleccionada:	SMT1	36	Los cálculos, un precio de la electricidad de tamaño medio de 19 céntimos / kwh.
Propuesta de otro motor:	SMT2	33	Los cálculos de los precios de la electricidad es un medio de 19 céntimos / kwh.
PE:			Los cálculos se basan en un precio medio de electricidad de 19 céntimos/kWh.

Tabla 251: Ejemplo de propuestas similares, con problemas de concordancia de número

Los casos en que se editaron errores en la traducción de los verbos separables no son tantos como para poder realizar conclusiones que tengan un valor representativo. Pero aquí, como las ediciones de los compuestos nominales vemos para el poseedor prefiere corregir a borrar adiciones o traducir de cero.

9.5 Sugerencias

Como presentamos ya en la introducción y en el capítulo 5 nuestro objetivo es establecer parámetros que puedan ayudar al poseedor y contribuir de esta manera en la formación académica de los traductores, que hoy en día no pueden dejar de lado el auge de los sistemas automáticos y deben ser preparados para poder ampliar su competencia como traductores

9. Incidencia de cada categoría de errores en el corpus de análisis

El análisis de los datos presentado a lo largo de este capítulo nos ha proporcionado una visión concreta sobre qué tipos de errores producen los motores y, ante diversas posibilidades, qué tipos de errores prefieren editar los poseedores. Toda esta información nos da pie a proponer algunos cambios en el planteamiento de los proyectos de posesición y su ejecución.

9.5.1 Sugerencias relativas a las fases anteriores a la posesición

A continuación enumeramos aquellos aspectos que podrían integrarse en la fase de preparación del proyecto, preparación del motor o tratamiento del texto traducido antes de su posesición humana, y que aumentarían:

1. La traducción incorrecta de nombres propios o extranjerismos podría evitarse o bien modificando el texto original o bien añadiendo reglas para el tratamiento de léxico específico, que bien pueden ser utilizadas en los motores basados en reglas o en sistemas híbridos de base estadística.
2. Errores detectados de alta incidencia, como por ejemplo el tratamiento incorrecto de las etiquetas, podría minimizarse durante la fase de preparación del texto original. Algo más de inversión en la revisión del texto original llevaría a un esfuerzo menor en la posesición, y esto especialmente relevante cuando se traduce en varias lenguas.
3. Errores de terminología podrían evitarse si el poseedor domina la temática o el motor ha sido entrenado con el diccionario terminológico correspondiente.
4. Los elementos que no se deben traducir, tales como comandos universales de programas —que se deben dejar por ejemplo en inglés— o nombres propios deben ser marcados como elementos que no se deben traducir. Es cuestionable si esto es o no posible en todos los tipos de motores de traducción automática, pero si es posible en la preparación del texto antes de la traducción.
5. En general los datos apuntan que una mayor inversión en la revisión del texto original llevaría a un esfuerzo menor en la posesición, y esto especialmente relevante cuando se traduce en varias lenguas. No se trataría únicamente de una revisión desde el punto de vista de la corrección original, sino también de identificar o adaptar aquellos elementos del texto que puedan suponer un problema de procesamiento para el motor de traducción y, en consecuencia, el origen de un error en el texto de salida.

9. Incidencia de cada categoría de errores en el corpus de análisis

6. Es importante que el poseedor esté familiarizado con el funcionamiento de la traducción automática, para que pueda asumir la tarea de posesición con una perspectiva diferente a la del corrector.

9.5.2 Sugerencias a las fases a realizar durante la posesición

Las sugerencias que damos a continuación pueden aplicarse a cualquier combinación lingüística, porque tienen más que ver con la tarea misma de la posesición, que con los problemas encontrados en ella. Este tipo de sugerencias se asemejan más a las directrices de posesición que enumeramos en el apartado.

1. Los poseedores deben poder tener la posibilidad de leer el texto original y no solo tener la salida o propuesta del motor de traducción automática.
2. Comprobar si el texto original está bien escrito.
3. Averiguar qué motor de traducción automática fue utilizado, por ejemplo, si se utilizó un sistema basado en reglas, ya que como vimos, los motores presentan problemas o errores específicos y diferenciados en función de la tecnología en la que se basan, y si el poseedor conoce el tipo de motor, podrá poner más atención a ciertos aspectos.
4. Determinar qué tipo de posesición se debe realizar y determinar qué expectativas de calidad se esperan.
5. Comprobar si se utilizó algún gestor de terminología y si no es así solicitar la terminología al cliente. En caso de que este no pueda darla, buscar ayudas terminológicas del área de conocimiento del encargo.

9.5.3 Sugerencias relativas a las fases posteriores a la posesición

En cuanto a la posesición entre alemán y castellano, el análisis realizado nos sugiere que, tanto en la formación de poseedores como en la elaboración de directrices cabría tener en cuenta que:

1. Los poseedores prefieren editar errores en palabras concretas que modificar la estructura de un segmento.
2. Los poseedores prefieren editar errores en palabras concretas que traducir palabras que se omitieron o borrar palabras que se añadieron.

9. Incidencia de cada categoría de errores en el corpus de análisis

3. En dos segmentos similares los poseedores prefieren editar errores de otro tipo a tener que ajustar adicionalmente errores de puntuación.
4. Como estrategia para la posesición se debería tener en cuenta que los motores de traducción automática no pueden reconocer siempre cuando una palabra es o no un nombre propio y cuándo debe o no traducirse.
5. Hay que considerar que en nuestra combinación lingüística la traducción de los compuestos nominales debe ser tenida en cuenta para la posesición, ya que se pueden dar muchos errores si estos no se interpretan correctamente.
6. En nuestra combinación lingüística hay que tener en cuenta que las formas verbales en infinitivo en alemán tienen la misma terminación que el imperativo impersonal que se utiliza en los textos técnicos para dar —por ejemplo— instrucciones. Este es un motivo frecuente de error en los textos del corpus.
7. En general, a partir de los datos obtenidos podemos afirmar que los errores de traducción habituales ente alemán y castellano descritos en el marco teórico son igualmente aplicables en el un entorno de posesición. De modo que más allá de qué aspectos prefieren editar los poseedores, los problemas de traducción que abordan en traducción y en posesición son esencialmente equiparables. En todo caso, la principal diferencia radica en situar la productividad como objetivo primordial en un entorno de posesición, así como partir de segmentos en lengua meta con errores de fluidez (gramaticales y sintácticos) que en un entorno TAO sería muy difícil encontrar.

El análisis de las ediciones realizadas durante la posesición nos dejó ver, que en la formación de los traductores se debe despertar una sensibilidad por la manera de ver los idiomas en cuestión. No solo se requiere tener la competencia lingüística, sino ser consciente de que ciertas características propias de la combinación particular de idiomas pueden ser problemáticas y generar errores, que en la traducción humana no se darían.

Por último, dado que todos son problemas propios de la traducción (ya sea provocados por la combinación lingüística, el tema, las características de texto original, las del texto final o el soporte en el que se publicará la traducción), parecía lógico que quien se ocupe de la gestión y la toma de decisiones en todas las fases del proyecto de traducción tenga un perfil de

9. Incidencia de cada categoría de errores en el corpus de análisis

experto en posesición, es decir un poseedor con un profundo conocimiento de las fases previas a la posesición. En este sentido la articulación de mecanismos para detectar cualquier acción repetitiva de los poseedores, causadas por las características del procesamiento del motor, del texto original o de la combinación lingüística permitiría aumentar la productividad del poseedor mediante la sistematización de soluciones.

9. Incidencia de cada categoría de errores en el corpus de análisis

PARTE IV: CONCLUSIONES

10. Conclusiones

El trabajo de posesición es todavía imprescindible si se emplean sistemas de traducción automática y no es en ningún momento un trabajo de menor importancia que el de la traducción sin este tipo de herramientas. La posesición es otro tipo de actividad —que es diferente de la corrección de textos traducidos por un humano— para la que el traductor debe realizar otro tipo de esfuerzo y desarrollar habilidades que debe tenerse en cuenta en la formación académica del traductor ya que hoy en día hacen parte de la competencia traductora.

En el marco teórico de la tesis encontramos que desde la perspectiva teórica existen ya aproximaciones a una definición amplia de los conceptos de traducción, de equivalencia y de finalidad misma de la traducción en las que el nuevo espectro de tareas del traductor como poseedor encaja perfectamente. También vimos que no se puede negar el hecho de que la traducción automática es una herramienta útil especialmente para la traducción de textos técnicos, que tienen una estructura similar, una función específica y un nivel de abstracción similar, que se diferencia, por ejemplo, de la traducción literaria o audiovisual.

También en el marco teórico vimos como la traducción de textos técnicos misma requiere que el traductor tenga conocimientos en el campo temático, deba utilizar la terminología correcta, y respetar el género del texto. El traductor de textos técnicos debe estar consciente de los problemas que se pueden presentar, tanto de orden general, como aquellos que tengan que ver con la combinación lingüística del encargo de traducción. Si el traductor tiene un trabajo de posesición de un texto es recomendable que tenga un conocimiento de los sistemas de traducción automática, de sus ventajas y sus deficiencias y de lo que se espera en su papel de poseedor. Recogemos aquí el concepto de “especificaciones” (Melby et al. 2014).

La aproximación teórica a las diferentes clasificaciones de problemas y errores de traducción nos ayudó en el análisis de los textos del corpus y también vimos los diferentes desarrollos de las tecnologías de traducción con énfasis en la traducción automática, sus diferentes métodos y sistemas. Al final del marco teórico presentamos el concepto de posesición, algunas de las directrices y reglas que se han desarrollado hasta el momento para esta tarea y para la evaluación misma del trabajo de posesición. Sin embargo creemos que estas reglas o directrices están más orientadas a ayudar en el ámbito comercial, es decir, a indicar la manera más efectiva y rápida de hacer la posesición y no vimos en ellas una reflexión acerca de cómo

10. Conclusiones

mejorar la formación académica de los traductores para que tengan la competencia necesaria para realizar esta tarea.

Las clasificaciones actuales relativas a las tipologías de errores en traducción están básicamente orientadas a tipificar los errores en la lengua de llegada, prescindiendo en gran medida de la lengua de partida. Incluso son aplicables a cualquier lengua de llegada, sin atender a los pormenores de cada lengua en concreto. La fortaleza de estas clasificaciones radica en que son aplicables en cualquier contexto. Sin embargo, solo permiten tipificar los errores de cada texto como si se tratase de errores aislados, sin contextualizarlos con la lengua de partida o la lengua de llegada, sin facilitar ver una sistematicidad en su origen y, por lo tanto, si dar pie a la generación de estrategias de resolución. Por ese motivo esas clasificaciones no son especialmente aptas para la introducción del traductor en el mundo de la posesición. Sin embargo, nuestra propuesta de clasificación sí que tiene en cuenta el contexto sistemático de cada error de traducción. De este modo se convierte en un instrumento formativo más rico y con más posibilidades, tanto para adquirir competencias específicas para la posesición como para detectar problemas que se den de manera sistemática y que deban repercutir en mejoras tecnológicas, incluso más allá de la tarea de posesición humana.

En el marco metodológico, después de presentar nuestros objetivos, hicimos un análisis preliminar de los datos de nuestro anotando los errores y comparándolos con las categorías de MQM. Este análisis tenía como objetivo el ayudarnos la mejor manera de procesar los datos y a partir de este análisis desarrollamos nuestra propuesta de clasificación de errores en cinco macrocategorías principales.

Con ayuda de esta clasificación evaluamos los datos del corpus y obtuvimos resultados que nos ayudaron a formular sugerencias específicas para la posesición de textos técnicos en nuestra combinación lingüística. Hasta este punto nuestra propuesta puede ser aplicada también en la industria de la traducción, pero nuestra reflexión estaba también encaminada a proponer estrategias que puedan ayudar a la formación académica de los traductores. Para ello realizamos en primer lugar una propuesta de clasificación de errores basada en la descripción de problemas, teniendo en cuenta cómo y cuándo se dan, con cuyo análisis pretendemos formular estrategias para la formación de traductores. Más allá que un conteo de errores, es una propuesta que pretende analizar qué preferencias tiene un poseedor analizando un caso concreto, con un tipo de texto concreto, en una combinación lingüística especial.

10. Conclusiones

La competencia tecnológica que debe tener un poseedor es diferente de la que ha tenido un traductor hasta el momento; pero con el aumento de la utilización de la traducción automática en combinación con la posesición tendrá que cambiar lo que entendemos por traducción y no solo en la reflexión teórica sino también en la práctica, es decir, cambiando la formación de los futuros especialistas en traducción, tanto si utilizan como si no utilizan traducción automática. De este modo, los resultados de esta tesis son una propuesta que se pueda convertir en un punto de partida para la formación académica de poseedores (Sánchez-Gijón 2016, 161).

En esta parte tendremos que volver a retomar otros de los aspectos mencionamos en el marco teórico. Nos referimos a los tipos y niveles de posesición (ver página 175). En guías para la posesición que presentamos en el marco teórico se diferenciaba el tipo de posesición a realizar, es decir, si deseábamos una posesición rápida para apenas entender de lo que se trata el texto, una posesición con una calidad ‘good enough’ o una posesición completa. Nuestra propuesta no está orientada en esta dirección.

Como nuestro objetivo es presentar criterios que ayuden a la formación de poseedores en el ámbito académico, el requisito primordial es tener conocimiento de la temática y de las distintas posibilidades. También con referencia al marco teórico tener una definición amplia de lo qué es la traducción, es decir, una definición en la que se incluya el tener conocimiento en la aplicación de traducción automática y de posesición en la competencia del traductor.

Por ello es importante la clasificación de macrocategorías y categorías de errores que hicimos, porque queremos sensibilizar al traductor en formación más que desarrollar una lista detallada de instrucciones para un traductor que reciba una tarea concreta de posesición. También son importantes los valores de los errores, lo que en otro tipo de clasificaciones se llaman tipos de errores, es decir si el error editado se dio porque se añadió, se omitió, se corrigió una palabra traducida erróneamente. Teniendo en cuenta estos valores pudimos detectar las preferencias de los poseedores en nuestro corpus.

En la industria de la traducción se plantea en ocasiones la pregunta de si la tarea de la posesición debe o tiene que ser realizada por un traductor o si solo se requiere que el poseedor tenga conocimientos en el lenguaje en que está escrito el texto que debe poseer. En esta

10. Conclusiones

tesis afirmábamos seguimos la idea de aquellos autores que afirman que proceso de la posesición debe ser realizado por un traductor, ya que solo un traductor puede juzgar la exactitud de la traducción ya que maneja también el lenguaje de origen, y por ello mismo está capacitado para corregir los errores en las propuestas de la traducción automática, gracias a los conocimientos que ha adquirido sobre la transferencia en varios idiomas (Kring 2001, 12).

La tarea de la traducción y la manera de trabajar del traductor se está transformando también en relación con el desarrollo tecnológico impulsado por la industria de la traducción. Por ello también creemos que se debe entender la traducción desde una definición amplia (Melby et al. 2014, 395) de esta que incluya la traducción automática y la posesición de traducciones automáticas como tareas que hacen parte de la competencia traductora.

Kring (2001) identifica una serie de categorías que nos ayudan a definir las tareas de la posesición como un compendio de procesos en las que se incluye en también la lectura del texto original; la lectura de la propuesta de traducción automática, y la producción de un texto meta. La mayor parte del proceso de posesición se realiza en esta última parte, ya que aquí que el poseedor se ocupa de producir un nuevo texto con todo lo que esto implica. Pero el poseedor puede además evaluar los resultados de la traducción automática y de esta manera ayudar al desarrollo tecnológico de los sistemas y motores de traducción automática. Esta retroalimentación puede estar destinada a mejorar los resultados de la traducción automática y/o mejor del contenido original, el hacer sugerencias de redacción o preedición, que puedan ayudar a resolver errores repetitivos en la salida de traducción automática. En este sentido los poseedores toman entonces el papel de actores en el escenario de la traducción (Rico y Torrejón 2012).

Recogemos la idea del artículo de Rico y Torrejón en que los autores exploran las nuevas habilidades o competencias que debe tener el traductor como poseedor de traducción automática que se agrupan en tres categorías principales: competencia básicas, habilidades lingüísticas y competencias instrumentales (Rico y Torrejón 2012, 169).

Las competencias básicas que están relacionadas con la capacidad que permite al poseedor, por ejemplo, manejar adecuadamente las expectativas del cliente en términos de aceptación de la calidad del texto. La competencia instrumental que ayuda a los poseedores a tomar decisiones informadas al elegir entre diferentes alternativas de edición, siguiendo las ins-

10. Conclusiones

trucciones o especificaciones, si es necesario, incluso aceptando una traducción de baja calidad.

Las habilidades lingüísticas que también son exigidas al traductor, y se refieren a tener el dominio del idioma de original y el tener una competencia comunicativa y textual en las lenguas y culturas implicadas, pero además el estar familiarizado con las reglas y las directrices de posesición.

Como competencia técnica se entiende también la habilidad técnica de entender qué es traducción automática, tener un conocimiento de los sistemas de traducción automática, sus capacidades y limitaciones. Saber gestionar terminología, diccionarios, habilidades de evaluación de la calidad del corpus. En algunos casos es ventajoso tener conocimientos en predicción de lenguaje controlado (Rico y Torrejón 2012), e incluso algunas habilidades de programación (para la creación de macros o scripts para la corrección o evaluación automatizada). El desarrollo de estas competencias debe tenerse en cuenta en la formación de traductores, en cualquier combinación lingüística.

En cuanto a la evaluación de los datos de nuestro corpus también podemos concluir que métodos como la distancia de Levenshtein—y algunas valoraciones de errores basados en este cálculo— no determina con exactitud la calidad de una salida de traducción automática y una buena puntuación no garantiza que el segmento sea favorecido, en un caso concreto como el nuestro, por un poseedor. Habría que buscar un nuevo instrumento o criterio de medición. Seguramente, lo que faltan aquí son métodos de investigación de tipo cognitivo. Ninguna medida automática nos podrá dar más información sobre por qué un poseedor elige un segmento u otro.

Con la aparición de los sistemas neuronales de traducción automática podría ser que algunos aspectos que se apuntan aquí como problemas desapareciesen, por ejemplo si ya no se dan tantos problemas y no hay que hacer tantas ediciones de orden sintáctico o corregir tantos errores de fluidez, pero también es posible que pueden aparecer otros problemas que habría que analizar. Pero sea cual fuese el motor o tipo de motor empleado, la tarea de la posesición es hoy en día parte de la tarea de traducción y por ello importante de considerar en la formación del traductor y en el desarrollo de la competencia traductora.

BIBLIOGRAFÍA

- Albrecht, Jörb. 2005. *Übersetzung und Linguistik*. Grundlagen der Übersetzungsforschung. Tübingen/Basel: Gunter Narr Verlag.
- Alcaraz Varó, Enrique. 2007. «Problemas metodológicos en la traducción del inglés jurídico». En *Problemas lingüísticos en la traducción especializada*, 17-32. Pedro A. Fuertes Olivera.
- Allen, Jeffrey. 2001. «Postediting: an integrated part of a translation software program». *Language International magazine* 13 (2): 26-29.
- . 2003. «Post-editing». En , editado por H. (.). Sommers, 297-317. *Computers and Translation: A Translator's Guide*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Aranberri, Nora. 2015. «SMT error analysis and mapping to syntactic, semantic and structural fixes». En *Proceedings of SSST-9*, 30-38. Denver, Colorado. <https://aclweb.org/anthology/W/W15/W15-1004.pdf>.
- Arnold, Douglas. 2003. «Why translation is difficult for computers». En *Computers and Translation: A translator's guide*, Inglés. Amsterdam/Philadelphia: Benjamins. <http://www.essex.ac.uk/linguistics/research/resgroups/clgroup/papers/doug/TranslatBook.pdf>.
- Arnold, Douglas, Lorna Balkan, Siety Meijer, Louisa Sadler, y R. Lee Humpreys. 1994. *Machine Translation: An Introductory Guide*. <http://promethee.philo.ulg.ac.be/engdep1/download/bacIII/Arnold%20et%20al%20Machine%20Translation.pdf>.
- Avramidis, Eleftherios. 2014. «Efforts on Machine Learning over Human-mediated Translation Edit Rate». En *Proceedings of the Ninth Workshop on Statistical Machine Translation*, 302–306. Baltimore. <http://www.statmt.org/wmt14/pdf/W14-3337.pdf>.
- Babych, Bogdan. 2014. «Automated MT evaluation metrics and their limitations». *Revista Tradumàtica: tecnologies de la traducció*, 464-70.
- Bader, Markus. 2004. «Syntax - Die Auflösung syntaktischer Ambiguitäten beim menschlichen Sprachverstehen». http://ling.uni-konstanz.de/pages/allgemein/study/introoling%20alt/einf_54.pdf.

Bibliografía

- Beck, Götz. 1973. «Textsorten und Soziolekte». En *Studien zur Theorie und zur der deutschen Grammatik*, 73-112. Düsseldorf.
- Belam, Judith. 2003. «Buying up to falling down. A deductive approach to teaching post-editing». En *MT Summit IX Workshop on Teaching Translation Technologies and Tools*. New Orleans.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.133.6685&rep=rep1&type=pdf>.
- Bermúdez, Montserrat. 2014. «Aplicación de las herramientas de traducción asistida por ordenador (TAO) a la investigación en traducción». *Estudios de Traducción*, n.º 4: 129-43.
- Bierwisch, Manfred. 1983. «Semantische und konzeptuelle Repräsentationen lexikalischer Einheiten». En *Untersuchungen zur Semantik*, editado por R. Ruzicka y W. Motsch, 61-99. Berlin: Akademie-Verlag.
- Blatz, John, Erin Fitzgerald, Georg Foster, Simona Gandrabur, Cyril Goutte, Alex Kulesza, Alberto Sanchis, y Nicola Ueffing. 2004. «Confidence Estimation for Machine Translation». En *Coling04*, 315-21. Geneva.
- Bowker, Lynne. 2002. *Computer-aided translation technology: A practical introduction*. Ottawa: University of Ottawa Press.
- Burchardt, Aljoscha, y Arle Lommel. 2014. «Practical Guidelines for the Use of MQM in Scientific Research on Translation Quality». <http://www.qt21.eu/downloads/MQM-usage-guidelines.pdf>.
- Callison-Burch, Chris, Philipp Koehn, Christof Monz, y Omar Zaidan. 2011. «Findings of the 2011 Workshop on Statistical Machine Translation.» En *Proceedings of the Sixth Workshop on Statistical Machine Translation (WMT 2011)*, 22-64. Edinburgh.
- Campbell, Stuart. 2000. «Choice Network Analysis in Translation Research». En *Intercultural Faultliness: Research Models in Translation Studies: Textual and Cognitive Aspects*, editado por M Olohan, 29-42. Manchester: St. Jerome Publishing.
- Carbonell, Jaime, Steve Klein, David Miller, Michael Steinbaum, Tomer Grassiany, y Joche Frey. 2006. «Context-Based Machine Translation». En , 19-28. Cambridge.
http://www.cs.cmu.edu/~jgc/publication/Context_Based_Machine_AMTA_2006.pdf.
- Carl, Michael, y Andy Way, eds. 2003. *Recent Advances in Example-Based Machine Translation*. Text, Speech and Language Technology 21. Kluwer Academic Publishers.
- Caro Cedillo, Ana. 2004. *Fachsprachliche Kollokationen. Ein übersetzungsorientiertes Datenbankmodell Deutsch-Spanisch*. Tübingen: Gunther Narr Verlag.
- Catford, John C. 1965. *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. London: Oxford University Press.
- . 1970. *Una teoría lingüística de la traducción; ensayo de lingüística aplicada*. Colección Avance 27. Caracas: Universidad Central de Venezuela.
- Cranias, Lambros, Harris Papageorgiou, y Stelios Piperidis. 1994. «Clustering: a technique for search space reduction in example-based machine translation». *IEEE International Conference on Systems, Man & Cybernetics* 1: 1-6.

Bibliografía

- Dalla-Zuanna, Jean-Marc. 2010. «Direkt Qualitätsmessung». *MDÜ Fachzeitschrift für Dolmetscher und Übersetzer* 2: 20-25.
- De Beaugrande, Robert A. 1981. «Introduction to Text Linguistics». En .
http://beaugrande.com/introduction_to_text_linguistics.htm.
- De Palma, Don. 2013. «tcworld.info - translation and localization».
<http://www.tcworld.info/e-magazine/translation-and-localization/article/post-editing-in-practice/>.
- Denkowski, Michael, y Alon Lavie. 2014. «Meteor Universal: Language Specific Translation Evaluation for Any Target Language». En *Proceedings of the EACL 2014 Workshop on Statistical Machine Translation*, 376--380. Baltimore: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W14/W14-3348>.
- DFKI GmbH, y QTLaunchPad. 2015. «Multidimensional Quality Metrics Issue Types».
<http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html>.
- «DFKI LT - taraXÜ». 2015. http://www.dfki.de/lt/project.php?id=Project_636&l=en.
- Diller, Hans-Jürgen, y Joachim Kornelius. 1978. *Lingüistische Probleme der Übersetzung*. Tübingen: Niemeyer.
- Dorr, Bonnie J. 1994. «Machine translation divergences: A formal description and proposed solution.» *Computational Linguistics* 20: 597-634.
- Eberle, Kurt, Ulrich Heid, Manuel Kountz, y Kerstin Eckart. 2008. «A tool for corpus analysis using partial disambiguation and bootstrapping of the lexicon» *Text Resources and Lexical Knowledge: Selected Papers from the 9th Conference on Natural Language Processing*.
- Eisele, Andreas. 2008. «Hybrid Architectures for Machine Translation». Wandlitz.
<http://www.statmt.org/mtm2/data/mtm2-eisele.pdf>.
- Elsen, Harald. 2012. «Maschinelle Übersetzung: Postediting – Schreckgespenst oder Perspektive». *MDÜ Fachzeitschrift für Dolmetscher und Übersetzer*, Die Zukunft des Übersetzens und Dolmetschens, 4: 16-17.
- Federmann, Christian. 2010. «Appraise: An Open-Source Toolkit for Manual Phrase-Based Evaluation of Translations». En *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010)*. Valletta, Malta.
<http://www.dfki.de/KI2012/PosterDemoTrack/ki2012pd27.pdf>.
- Flórez, Silvia. 2012. «Catálogo de herramientas para la posesición». *Revista Tradumàtica: tecnologies de la traducció* 10 (Postedició, canvi de paradigma?): 219-29.
- Forcada, Mike, Mireia Ginestí-Rosell, y Jacob Nordfalk. 2011. «Apertium: a free/open-source platform for rule-based machine translation» 25 (2): 127-44.
- Fränkel, Caroline. 2008. «Evaluation von Syntaxanalysemethoden angewendet auf Datenmaterial aus transliterierten Schulaufsätzen». <https://kola.opus.hbz-nrw.de/frontdoor/index/index/year/2008/docId/225>.
- Gamero, Silvia. 1998. *La traducción de textos técnicos (alemán-español): Géneros y subgéneros*. Universitat Autònoma de Barcelona.

Bibliografía

- . 2001. *La traducción de textos técnicos*. Editorial Ariel, S.a.
- . 2010. *Traducción alemán-español: aprendizaje activo de destrezas básicas*. Universidad Jaume I.
- García Yebra, Valentín. 1984. «La interferencia Lingüística». En *Teoría y práctica de la traducción*, 2a. Edición, I:353-84. Biblioteca Románica Hispánica, III. Madrid: Gredos.
- Gentzler, Edwin. 1993. *Contemporary Translation Theories*. Routledge.
- GitHub. 2014. «Appraise». <https://github.com/cfedermann/Appraise>.
- Görög, Attila. 2014. «Quantificació i avaluació comparativa de la qualitat: el Dynamic Quality Framework de TAUS». *Revista Tradumàtica: tecnologies de la traducció* 12: 443-54.
- Görz, Günther, y Josef Schneeberger, eds. 2003. *Handbuch der künstliche Intelligenz*. Oldenburg: Oldenburg Wissenschaftsverlag GmbH.
- Gouadec, Daniel. 2007. «Translation as a Profession». En , 396. Amsterdam, Philadelphia: John Benjamins B.V.
- Green, Roy. 1982. «The MT Errors which Cause Most Trouble to Post-Editors». En *Practical Experience of Machine Translation: Proceedings of a Conference*, 101-4. <http://www.mt-archive.info/70/Aslib-1981-Green.pdf>.
- Groves, Declan, y Dag Schmidtke,. 2009. «Identification and analysis of post-editing patterns for MT». En *Proceedings of MT Summit XII*, 429-36. Ottawa, Ontario, Canada. <http://www.mt-archive.info/MTS-2009-Groves.pdf>.
- Guerberof Arenas, Ana. 2008. «Productivity and quality in the post-editing of outputs from translation memories and machine translation». *Localisation Focus: The International Journal of Localisation* 7 (1): 11-21.
- Hatim, Basil, y Ian Mason. 1990. *Discourse and the Translator*. London, New York: Longman.
- Hermans, Theo. 1991. «Translational Norms and correct Translations». En *Translation Studies: The State of the Art: Proceedings of the First James S. Holmes Symposium on Translation Studies*, editado por Kitty M van Leuven-Zwart y Ton Naaijken, 155-70. Amsterdam: Rodopi.
- Hoffmann, Lothar. 1987. *Kommunikationsmittel Fachsprach. Eine Einführung*. Berlin: Akademie-Verlag.
- House, Juliane. 1977. *A Model for Translation Quality Assessment*. Tübingen: Gunter Narr Verlag.
- Hurtado, Amparo. 2011. *Traducción y Traductología: Introducción a la traductología*. Sexta. Madrid: Ediciones Cátedra (Grupo Anaya, S.A.).
- Hutchins, W. John. 2001. «Machine translation over fifty years». *Histoire, Epistemologie, Langage* XXII (1): 7-31.
- Hutchins, W. John, y Harold L. Somers. 1992. *An introduction to machine translation*. Londres: Academic Press.

Bibliografía

- Jäger, Gert. 1975. *Translations und Translationslinguistik*. Linguistische Studien. Halle: Niemeyer,.
- Jakobsen, Arnt Lykke. 1998. «Logging time delay in translation». *Copenhagen Working Papers in LSP* 1: 73-101.
- Jakobson, Roman. 1959. «On Linguistic Aspects of Translation». En *On Translation*, editado por R.A. Brower, 232-39. Harvard University Press.
- Jumplet, Rudolf W. 1961. *Die Übersetzung naturwissenschaftlicher und technischer Literatur*. Berlin: Langenscheidt.
- Jurafsky, Daniel, y James H. Martin. 2006. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Segunda. Upper Saddle River : Pearson Education (US).
<http://plaza.ufl.edu/hfilip/jurafsky.martin24.pdf>.
- Kade, Otto. 1968. *Zufall und Gestetzmässigkeit in der Übersetzung*. Leipzig: VEB Verlag Enzyklopädie.
- Kelly, Dorothy A. 2002. «Un modelo de competencia traductora: bases para el diseño curricular». *Puentes*, n.º 1 (Enero): 9-20.
- Koehn, Philipp. 2009. «A process study of computer-aided translation». *Machine Translation* 23 (4): 241-63.
- . 2010. *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- Koller, Werner. 2011. *Einführung in die Übersetzungswissenschaft*. Wibelsheim: Quelle & Mexer Verlag.
- Königs, Frank G. 1981. «Zur Frage der Übersetzungseinheit und ihre Relevanz für den Fremdsprachenunterricht». *Linguistische Berichte* 74: 82-103.
- Krings, Hans P. 1986. *Was in den Köpfen von Übersetzern vorgeht*. Tübinger Beiträge zur Linguistik. Tübingen: Narr Verlag.
- . 2001. *Repairing Texts. Empirical Investigations of Machine Translation Post-Editing Processes*. Editado por G.S Boby. Kent: The Kent State University Press.
- . 2005. «Wege ins Labyrinth – Fragestellungen und Methoden der Übersetzungsprozessforschung im Überblick». *META: journal des traducteurs / Meta: Translators' Journal* 50 (2): 342-58.
- Lavanchy, Claire. 2010. «Statistische Maschinelle Übersetzung». <http://kitt.ifi.uzh.ch/clab/smt/smt.pdf>.
- Lavorel, Bernard. 1982. «Experience in English-French Post-Editing». En *Practical Experience of Machine Translation: Proceedings of a Conference*, editado por V. Lawson, 105-9. London: North-Holland Publishing Company. <http://www.mt-archive.info/70/Aslib-1981-Lavorel.pdf>.
- Löbner, Sebastian. 2003. *Semantik*. Berlin: De Gruyter.
- Loffler-Laurian, Anne Marie. 1983. «Pour une typologie des erreurs dans la traduction automatique». *Multilingua - Journal of Cross-Cultural and Interlanguage Communication* 2 (2): 65-78.

Bibliografía

- . 1984. «Machine Translation : What Type of Post-Editing on What Type of Documents for What Type of Users». En , 236-38. Stroudsburg, PA, USA. doi:10.3115/980491.980542.
- Lommel, Ariel, Aljoscha Burchardt, Attila Görög, Hans Uszkoreit, y Alan Melby. 2016. «Multidimensional Quality Metrics (MQM) Issue Types». *Multidimensional Quality Metrics (MQM) Issue Types*. septiembre 2. <http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html>.
- Lommel, Ariel, Hans Uszkoreit, y Aijoscha Burchardt. 2014. «Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics». *Revista Tradumàtica: tecnologies de la traducció* 12: 455-63.
- «Machine Translation Evaluation for GALE». 2006. <http://www.nist.gov/itl/iad/mig/gale.cfm>.
- Maegaard, Bente. 2001. «Machine Translation.» En , editado por Ed Hovy, Robert Frederking, y Antonio Zampolli, 81-103. *Multilingual Information Management: Current Levels and Future Abilities*. Pisa-Roma, Italia: Istituti editoriale e poligrafici internazionali.
- Martin-Mor, A. 2017. «MTradumàtica: Statistical Machine Translation Customisation for Translators». *Skasse Journal of Translation and Interpretation* 11 (1). http://ddd.uab.cat/pub/artpub/2017/174544/skase_published.pdf.
- Martin-Mor, Adrià. 2011. «La interferència lingüística en entorns de Traducció Assistida per Ordinador». Tesis doctoral, Barcelona: Universidad Autónoma de Barcelona. http://ddd.uab.cat/pub/tesis/2011/hdl_10803_83987/amm1de1.pdf.
- Martin-Mor, Adrià, Ramón Pique, y Pilar Sánchez-Gijón. 2016. *Tradumàtica Tecnologies de la traducció*. Romanyà Valls: Eumo Editorial.
- Martín-Mor, Adria, Ramón Pique-Huerta, y Pilar E. Sánchez-Gijón. 2016. «Cambios en el paradigma de la traducción digitalizada».
- McElhaney, Terrence, y Muriel Vasconcellos. 1988. «The Translator and the Postediting Experience». *Technology as translation strategy* 2: 140-48.
- Melby, Alan, Paul Fields, Daryl Hague, Geoffrey S. Koby, y Arle Lommel. 2014. «Defining the Landscape of Translation». *Revista Tradumàtica: tecnologies de la traducció*, n.º 12: 392-403.
- Mellado Blanco, Carmen. 1999. «La metáfora en la formación de fraseologismos alemanes y españoles: las metáforas locales». *Paremia* 8: 333-38.
- . 2009. «La pupila es la ‘niña’: las metá foras de los lexemas somáticos del alemán y el español». *Paremia* 18: 53-63.
- Meng, Michael, y Markus Bader. 2000. «Mode of Disambiguation and Garden-Path Strength: An Investigation of Subject-Object Ambiguities in German». *Language & Speech* 43 (1): 43-74.
- Microsoft. 2016. «Microsoft Translator launching Neural Network based translations for all its speech languages». noviembre. <https://blogs.msdn.microsoft.com/translation/2016/11/15/microsoft-translator-launching-neural-network-based-translations-for-all-its-speech-languages/>.

Bibliografía

- Mittag, Marc. 2013. *translate5*. Mössingen: MittagQI - Quality Informatics.
- «MQM (Multidimensional Quality Metrics)». 2014. <http://www.qt21.eu/downloads/MQM-overview-2014-08-14.pdf>.
- Navarro, Fernando A. 1996. «Glosario de “falsos amigos” en el alemán de los textos médicos». *Lebende Sprachen* 4: 181-87.
- Neubert, Albrecht. 1985. *Text and Translation*. Leipzig: VEB Verlag Enzyklopädie.
- Nida, Eugene A. 1959. «Principles of Translation as exemplified by Bible Translating». En *On Translation*, editado por R.A. Brower, 11-13. Harvard University Press.
- Nida, Eugene A. 1964. *Toward a Science of Translating with Special Reference to Principles and Procedures involved in Bible Translating*. Leiden: E. J. Brill.
- Nida, Eugene A., y Charles Russel Taber. 1969. *Theorie und Praxis des Übersetzens unter besonderer Berücksichtigung der Bibelübersetzung*. Weltbund der Bibelgesellschaften.
- Nissen, Sonja, Franz J. Och, Greogor Leusch, y Hermann Ney. 2000. «An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research». En *Proceedings of the 2nd International Conference on Language Resources and Evaluation*. Atenas, Grecia.
- NIST. 2015. «Open Machine Translation Evaluation». <https://www.nist.gov/itl/iad/mig/open-machine-translation-evaluation>.
- Nord, Christiane. 1987. «Übersetzungsprobleme-Übersetzungsschwierigkeiten. Was in den Köpfen von Übersetzern vorgehen sollte». *Mitteilungsblatt für Dolmetscher und Übersetzer* 2: 5-8.
- . 1991. *Textanalyse und Übersetzen*. 2. neu bearbeitete Auflage. Heidelberg: Julius Gross Verlag.
- . 1996. «El error en la traducción: categorías y evaluación». En *Estudis sobre la traducció*, editado por Amparo Hurtado, 91-107. Castellón: Universitat Jaume I.
- . 2001a. *Lernziel: Professionelles Übersetzen Spanisch-Deutsch*. Gottfried Egert Verlag.
- . 2001b. *Lernziel: Professionelles Übersetzen Spanisch-Deutsch*. Gottfried Egert Verlag.
- . 2009. «El funcionalismo en la enseñanza de la traducción». *Mutatis Mutandis* 2 (2): 209-43.
- O'Brien, Sharon. 2005. «Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability - Springer». *Journal Machine Translation*, marzo.
- . 2006. «Machine-Translatability and Post-Editing Effort: An Empirical Study using Translog and Choice Network Analysis». Dublin: Dublin City University.
- . 2007. «Eye-tracking and Translation Memory Matches». *Perspectives: Studies in Translatology* 14 (3): 185-205.

Bibliografía

- . 2010. «Introduction to Post-Editing: Who, What, How and Where to Next». presentado en AMTA 2010, Denver, Colorado. <http://www.mt-archive.info/AMTA-2010-O'Brien.pdf>.
- . 2012. «Towards a dynamic quality evaluation model for translation». *JoSTrans*, n.º Issue 17: 55-77.
- Oettinger, Anthony G. 1960. *Automatic Language Translation*. Harvard University Press.
- «OpenOffice». 2009. <http://opus.lingfil.uu.se/OpenOffice.php>.
- Ortiz Jimenez, María. 2014. «Traducción asistida y traductores autónomos: alternativas y avances». http://ec.europa.eu/translation/bulletins/puntoycoma/123/pyc1237_es.htm.
- Oster, Ulrike. 2003. «Los términos de la cerámica en alemán y en español. Análisis semántico orientado a la traducción de los compuestos nominales alemanes». Castellón: Universitat Jaume I. <http://www.tdx.cat/bitstream/handle/10803/10565/oster.pdf;jsessionid=E22C21FEB309E1B380E64C21CF560305?sequence=1>.
- Papineni, Kishore, Salim Roukos, Todd Ward, y Wei-Jing Zhu. 2002. «Bleu: a method for automatic evaluation of machine translation». En *Proceedings of the 40th annual meeting on association for computational linguistics*, 311-18. Philadelphia, Pennsylvania. <http://www.aclweb.org/anthology/P02-1040.pdf>.
- Pinchuck, Isadore. 1977. *Scientific and Technical Translation*. London: Andre Deutch.
- Popovic, Maja, Eleftherios Avramidis, Aljoscha Burchardt, David Vilar, y Hans Uszkoreit. 2013. «What can we learn about the selection mechanism for post-editing?» En *Proceedings of Workshop on Translation Post-Editing Technology and Practice. Machine Translation Summit (MT-Summit-13)*. Xiamen, China. http://www.dfki.de/web/research/publications/renameFileForDownload?filename=finalSubmission.pdf&file_id=uploads_2191.
- Pym, Anthony. 2012. *Teorías contemporáneas de la traducción*. Taragona. http://isg.urv.es/publicity/isg/publications/2011_teorias/index.htm.
- Ramlow, Markus. 2008. «Maschinelle Übersetzungssysteme im Vergleich». En *Maschinelle Übersetzung und XML im Übersetzungsprozess*. Vol. 19. Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens. Berlin: Frank & Timme GmbH.
- . 2009. *Die maschinelle Simulierbarkeit des Humanübersetzens*. Berlin: Frank & Timme.
- Reimann, Martina. 1991. *Maschinelle Übersetzung: Ein Überblick über Theorie und Praxis*. Berlin: Springer-Verlag.
- Reinke, Uwe. 2003. *Translation Memories: Systeme - Konzepte - Linguistische Optimierung*. Peter Lang Verlag.
- Reischer, Jürgen. 2002. *Die Sprache*. Berlin: De Gruyter.
- Reiss, Katharina, y Hans J. Vermeer. 1996. *Fundamentos para una teoría funcional de la traducción*. Madrid: Ediciones Akal.
- «Reverso». 2016. Comercial. *Softissimo*. <http://www.softissimo.com/reverso.asp>.

Bibliografía

- Rico, Celia, y Enrique Torrejón. 2012. «Skills and Profile of the New Role of the Translator as MT Post-editor». *Revista Tradumàtica*, n.º 10: 166-78.
- Rodríguez Díaz, José Antonio. 2011. «Anglicismos y germanismos en el lenguaje de la publicidad del español peninsular contemporáneo (1998-2007)». Santiago de Compostela: Universidad de Santiago de Compostela.
http://minerva.usc.es/bitstream/10347/3636/1/9788498877977_content.pdf.
- SAE International. 2005. «SAE J2450 Translation Quality Metric Task Force».
<http://www.sae.org/standardsdev/j2450p1.htm>.
- Sánchez, Alfonso Corbaco. 2004. «Breve estudio sobre los germanismos del alemán en el lenguaje de la economía». *Anuario de estudios filológicos*, 27 edición.
- Sánchez-Gijón, Pilar. 2016. «La posesición: hacia una definición competencial del perfil y una descripción multidimensional del fenómeno». *Sendebarr*, n.º No. 27: 151-62.
- Sánchez-Martínez, Felipe. 2012. «Motivos del creciente uso de traducción automática seguida de posesición». *tradumàtica*, n.º 10: 150-56.
- . 2014. «Statistical Machine Translation for Translators». Presentación Power Point presentado en Escuela de Verano UAB 2014, Barcelona.
- Schmitt, Peter A. 1999. *Translation und Technik*. Tübingen: Stauffenburg.
- Schumann, Ingo, y Mathan Sivaloganathan. 2013. «Post-Editing maschineller Übersetzung in der Praxis». presentado en Tekom Herbst Tagung, Wiesbaden.
http://www.lucyd.org/material/Post-EditingTekom2012is_ms.pdf.
- Searle, John. 1969. *Actos de habla: Ensayo de filosofía del lenguaje*. Teorema Serie Mayor. Buenos Aires: Planeta Agostini. <http://www.textosenlinea.com.ar/libros/Searle%20-%20Actos%20de%20Habla.pdf>.
- Seleskovitch, Danica, y Marianne Lederer. 1984. «Intérpreter pour traduire». *L'Information Grammaticale*. http://www.persee.fr/web/revues/home/prescript/article/igram_0222-9838_1985_num_25_1_2204_t1_0044_0000_4#.
- Sevilla Muñoz, Manuel, y Julia Sevilla Muñoz. 2003. «Una clasificación del texto científico-técnico desde un enfoque multidireccional». *Language Design* 5: 19-38.
- Snell-Hornby, Mary. 1988. *Translation Studies: An Integrated Approach*. Amsterdam: John Benjamins Publishing Company.
- Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, y John Makhoul. 2006. «A Study of Translation Edit Rate with Targeted Human Annotation». En *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*, 223-31. Cambridge. <http://mt-archive.info/AMTA-2006-Snover.pdf>.
- Somers, H., ed. 2003a. *Computers and translation: A translator's guide*. Amsterdam, Philadelphia: John Benjamins.
- Somers, Harold. 2003b. «An Overview of EBMT». En *Recent advances in Example-Based Machine Translation*, editado por Michael Carl y Way, 3-57. Dordrecht: Springer Netherlands.

Bibliografía

- Specia, Lucia. 2013. «Translation Quality Assessment: Evaluation and Estimation». En . University of Sheffield. <http://de.slideshare.net/RIILP/10-lucia-specia-usfd-evaluation-of-machine-translation>.
- Specia, Lucia, y Atefeh Farzindar. 2010. «Estimating Machine Translation Post-Editing Effort with HTER». En . The Ninth Conference of the Association for Machine Translation in the Americas. http://retour.iro.umontreal.ca/rali/sites/default/files/publis/Specia-Farzindar_AMTA_workshop.pdf.
- Specia, Lucia, Kashif Shah, Jose G.C. De Sousa, y Trevor Cohn. 2013. «QuEst - A translation quality estimation framework». En *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 79-84. Sofia, Bulgaria. <http://www.aclweb.org/anthology/P13-4014>.
- Stein, Daniel. 2009. «Maschinelle Übersetzung - ein Überblick». *JLCL* 24 (3): 5-18.
- Stolze, Radegundis. 2011. *Übersetzungstheorien: Eine Einführung*. 6.^a ed. Tübingen: Gunter Narr Verlag.
- Tatsumi, Midori. 2010. «Post-Editing Machine Translated Text in A Commercial Setting: Observation and Statistical Analysis». Thesis submitted for the degree of Doctor of Philosophy, Dublin: Dublin City University. http://doras.dcu.ie/16062/1/SAKURA_final_revised.pdf.
- TAUS. 2014. «Adequacy/Fluency Guidelines». <https://www.taus.net/think-tank/best-practices/evaluate-best-practices/adequacy-fluency-guidelines?ModPagespeed=noscript>.
- . 2015. «Rules of Post-editing - knowledgebase». https://www.taus.net/knowledgebase/index.php?title=Rules_of_Post-editing.
- TAUS/CNGL. 2011. «Machine Translation Post-editing Guidelines». <http://www.cngl.ie/node/2542>.
- Tiedemann, Jorg. 2009. «News from OPUS – A Collection of Multilingual Parallel Corpora with Tools and Interfaces». En *Recent Advances in Natural Language Processing*, V:237-48. Borovets.
- Tomás, Jesús. 2003. «Traducción automática de textos entre lenguas similares utilizando métodos estadísticos». Valencia: Universidad Politécnica de Valencia. <http://personales.upv.es/~jtomas/articulos/tesis.pdf>.
- Uszkoreit, H. 2002. «Einführung in die Computerlinguistik Teil 2». Universität des Saarlandes. <http://slideplayer.org/slide/646128/>.
- Uszkoreit, Hans. 1998. «Einführung in die Computerlinguistik». Uni Saarland. <http://www.coli.uni-saarland.de/~hansu/VLCL1/VLCLTeil1.pdf>.
- . 2014. «QTLaunchPad». diciembre 5. <http://www.qt21.eu/launchpad/>.
- Van Gompel, Maarten. 2009. «Phrase-based Memory-based Machine Translation». Tilburg University.
- Vannerem, Mia, y Mary Snell-Hornby. 1986. «Die Szene hinter dem Text: “scenes-and-frames semantics” in der Übersetzung.» En *Übersetzungswissenschaft. Eine Neuorientierung*, 184-205. Tübingen: Francke.

Bibliografía

- Vásquez-Ayora, Gerardo. 1977. *Introducción a la traductología: curso básico de traducción*. Georgetown University Press.
- Vauquois, Bernard. 1975. «La traduction automatique à Grenoble». *Document de Linguistique Quantitative*, n.º 24: 180.
- Veale, Tony, y Andy Way. 1997. «Gaijin: A template-based bootstrapping approach to example-based machine translation». En *Proceedings of "New Methods in Natural Language Processing"*. Sofía.
- Vilar, David, Jia Xu, Luis F. D'Haro, y Hermann Ney. 2006. «Error Analysis of Statistical Machine Translation Output». En *International Conference on Language Resources and Evaluation*, 697-702. Génova. <http://www-i6.informatik.rwth-aachen.de/publications/downloader.php?id=281&row=pdf>.
- Vinay, Jean-Paul, y Jean Darbelnet. 1958. *Stylistique comparée du français et de l'anglais: méthode de traduction*. Paris: Didier.
- Wagner, Emma. 1985. «Post-Editing Systran – A Challenge for Commission Translator». *Terminologie et Traduction*. <http://mt-archive.info/T&T-1985-Wagner.pdf>.
- . 1987. «Post-Editing: Practical Considerations». En *Proceedings of ITI Conference. I: The Business of Translating and Interpreting*, editado por C. Picken, 71-78. London: Aslib.
- Wills, Wolfram. 1977. *Übersetzungswissenschaft: Probleme und Methoden*. Stuttgart: Klett-Cotta.
- Winter, Werner. 1961. «Impossibilities of Translation». En *The Craft and Context of Translation*, 62-82. Ed. William Arrowsmith and Roger Shattuck. Austin Texas: University of Texas.

Bibliografía

ÍNDICE DE ILUSTRACIONES Y TABLAS

Índice de ilustraciones

Ilustración 1:	Grados de automatización del proceso de traducción según Hutchins y Somers(1992, 148)	74
Ilustración 2:	Fases y tareas del proceso del proceso de digitalización de la traducción (Adrià Martin-Mor, Pique, y Sánchez-Gijón 2016, 34)	75
Ilustración 3:	Modelos de traducción automática (Lavanchy 2010, 2).....	80
Ilustración 4:	Modelo basado en la interlingua con seis pares de lenguas (Hutchins y Somers 1992, 74).....	84
Ilustración 5:	Triangulo de Vauquois (Van Gompel 2009, 16)	85
Ilustración 6:	Alienación según Koehn (2010, 56).....	86
Ilustración 7:	Formula de Bayes (Koehn 2010, 69).....	87
Ilustración 8:	Comparación de un corpus paralelo (Koehn 2010, 183)	89
Ilustración 9:	Marcaciones de combinaciones en SMT (Koehn 2010, 8).....	91
Ilustración 10:	Tipo de ordenamiento monótono (Sánchez-Martínez 2014).....	92
Ilustración 11:	Tipo de ordenamiento discontinuo (Sánchez-Martínez 2014).....	92
Ilustración 12:	Ponderación léxica (Koehn 2010, 140)	93
Ilustración 13:	Diferencias en la posición del verbo en alemán y castellano	123
Ilustración 14:	Comparación de errores presentados en SMT	124
Ilustración 15:	Hoja de resultados de un proyecto evaluado con SA-J2450 (Dalla-Zuanna 2010, 23).....	135
Ilustración 16:	Algoritmo de definición de BLEU (Sánchez-Martínez 2014).....	142
Ilustración 17:	Categorías de errores de MQM (Lommel, Burchardt, Görög, Uszkoreit & Melby, 2016)	154
Ilustración 18:	Jerarquía de los errores en la categoría Accuracy (Lommel et al. 2016).....	155
Ilustración 19:	Jerarquía de los errores en la categoría Design (Lommel et al. 2016).....	157
Ilustración 20:	Jerarquía de los tipos de errores en la categoría Fluency (Lommel et al. 2016) ...	159
Ilustración 21:	Jerarquía de la categoría de errores <i>Verity</i> (Lommel et al. 2016).....	162

Índice de ilustraciones

Ilustración 22:	Jerarquía de la categoría de errores <i>Terminology</i> (Lommel et al. 2016)	163
Ilustración 23:	Jerarquía de la categoría de errores <i>Style</i> (Lommel et al. 2016).....	164
Ilustración 24:	Jerarquía de la categoría de errores <i>Locale convention</i> (Lommel et al. 2016)	165
Ilustración 25:	Jerarquía de la categoría de errores <i>Internationalization</i> (Lommel et al. 2016)....	167
Ilustración 26:	Esquema de relación de MQM («MQM (Multidimensional Quality Mectrics)» 2014).....	209
Ilustración 27:	Tabla de datos con conteo de palabras y de distancia de teclado	221
Ilustración 28:	Ejemplo de un árbol de decisiones	227
Ilustración 29:	Interface de usuario de translate5	228
Ilustración 30:	Tabla de conteo de errores según MQM	229
Ilustración 31:	Errores encontrados en las categorías generales de MQM.....	230
Ilustración 32:	Tabla para la anotación de los errores	233
Ilustración 33:	Segmentos poseídos a partir de cada sistema (incluyendo las propuestas de la TM).....	234
Ilustración 34:	Hoja de cálculo para conteo de errores según motor y tipo.....	234
Ilustración 35:	Porcentaje de errores según motor empleado para la traducción	235
Ilustración 36:	Propuestas marcadas como correctas según el motor.....	236
Ilustración 37:	Problemas/errores marcados en los segmentos originales.....	238
Ilustración 38:	Porcentaje de errores más frecuentes independientemente del motor de traducción empleado.....	254
Ilustración 39:	Errores más frecuentes en las propuestas de SMT1	256
Ilustración 40:	Errores más frecuentes en las propuestas de SMT2	258
Ilustración 41:	Errores más frecuentes en las propuestas del motor RBMT1.....	260
Ilustración 42:	Errores más frecuentes en las propuestas de RBMT2	262
Ilustración 43:	Porcentaje de traducciones según el valor de la distancia de teclado.....	268
Ilustración 44:	Errores más frecuentes encontrados en la categoría <i>Accuracy</i>	270
Ilustración 45:	Errores encontrados en la subcategoría <i>Mistranslation</i>	271
Ilustración 46:	Errores encontrados en la categoría <i>Fluency</i>	275
Ilustración 47:	Tabla para el conteo de errores den la segunda fase.....	311
Ilustración 48:	Porcentaje de propuestas marcadas como correctas según el motor	311
Ilustración 49:	Porcentaje de propuestas seleccionadas para la posesición según el motor.....	312
Ilustración 50:	Número de errores encontrados en relación con los motores seleccionados.....	312
Ilustración 51:	Porcentaje de errores únicos más frecuentes sin tener en cuenta el motor	313
Ilustración 52:	Porcentaje de errores únicos más frecuentes en las propuestas del motor rbmt1 ..	314
Ilustración 53:	Errores más frecuentes independientemente del motor seleccionado	316
Ilustración 54:	porcentaje de errores marcados según las macrocategorías	317

Índice de ilustraciones

Ilustración 55:	Errores más frecuentes en las propuestas de SMT1	318
Ilustración 56:	Porcentaje de errores encontrados en las propuestas del smt1 según las macrocategorías definidas	320
Ilustración 57:	Errores más frecuentes en las propuestas de SMT2	321
Ilustración 58:	Porcentaje de errores encontrados en las propuestas del smt2 según las macrocategorías definidas	323
Ilustración 59:	Errores más frecuentes en las propuestas de RBMT1	324
Ilustración 60:	Porcentaje de errores encontrados en las propuestas del rbmt1 según las macrocategorías definidas	326
Ilustración 61:	Errores más frecuentes en las propuestas de RBMT2	326
Ilustración 62:	Porcentaje de errores encontrados en las propuestas del rbmt2 según las macrocategorías definidas	328
Ilustración 63:	Comparación de los errores más frecuentes según el motor seleccionado	334
Ilustración 64:	Porcentaje de errores encontrados según las macrocategorías	340
Ilustración 65:	Porcentaje total de errores marcados en la primera macrocategoría	341
Ilustración 66:	Porcentaje total de errores marcados en la segunda macrocategoría.....	343
Ilustración 67:	Porcentaje total de errores marcados en la tercera macrocategoría	344
Ilustración 68:	Porcentaje total de errores marcados en la cuarta macrocategoría en palabras sin carga semántica	345
Ilustración 69:	Porcentaje total de errores marcados en la cuarta macrocategoría en palabras con carga semántica.....	347
Ilustración 70:	Porcentaje total de errores marcados en la quinta categoría.....	348
Ilustración 71:	Porcentaje total de errores marcados en la sexta macrocategoría.....	349
Ilustración 72:	Número de errores marcados clasificados por macrocategorías según el motor seleccionado	351
Ilustración 73:	Incidencia de los tipos de motores en las categorías de la primera macrocategoría	352
Ilustración 74:	Errores encontrados en la primera macrocategoría SMT	352
Ilustración 75:	Errores encontrados en la primera macrocategoría RBMT	353
Ilustración 76:	Incidencia de los tipos de motores en las categorías de la segunda macrocategoría	353
Ilustración 77:	Errores encontrados en la segunda macrocategoría SMT.....	354
Ilustración 78:	Errores encontrados en la segunda macrocategoría RBMT	354
Ilustración 79:	Incidencia de los tipos de motores en las categorías de la tercera macrocategoría.....	355
Ilustración 80:	Errores encontrados en la tercera macrocategoría SMT	355
Ilustración 81:	Errores encontrados en la tercera macrocategoría RBMT.....	356
Ilustración 82:	Incidencia de los tipos de motores en las categorías de la cuarta macrocategoría sin carga semántica.....	356

Índice de ilustraciones

Ilustración 83:	Incidencia de los tipos de motores en las categorías de la cuarta macrocategoría con carga semántica.....	357
Ilustración 84:	Errores encontrados en la cuarta macrocategoría SMT	357
Ilustración 85:	Errores encontrados en la cuarta macrocategoría RBMT	358
Ilustración 86:	Incidencia de los tipos de motores en las categorías de la quinta macrocategoría	358
Ilustración 87:	Errores encontrados en la quinta macrocategoría SMT.....	359
Ilustración 88:	Errores encontrados en la quinta macrocategoría RBMT.....	359
Ilustración 89:	Relación ponderada de los errores encontrados en cada macrocategoría y en cada motor	360
Ilustración 90:	Relación por motor en las categorías de la primera macrocategoría	361
Ilustración 91:	Relación del motor SMT1 en cada categoría de la primera macrocategoría	362
Ilustración 92:	Relación del motor SMT2 en cada categoría de la primera macrocategoría	363
Ilustración 93:	Relación del motor RBMT1 en cada categoría de la primera macrocategoría	364
Ilustración 94:	Relación del motor RBMT2 en cada categoría de la primera macrocategoría	365
Ilustración 95:	Relación tipo de motor categorías de la segunda macrocategoría	367
Ilustración 96:	Relación del motor SMT1 en cada categoría de la segunda macrocategoría	368
Ilustración 97:	Relación del motor RBMT1 en cada categoría de la segunda macrocategoría	369
Ilustración 98:	Relación por motor y categoría en la tercera macrocategoría	371
Ilustración 99:	Relación del motor SMT1 en cada categoría de la tercera macrocategoría.....	372
Ilustración 100:	Relación del motor SMT2 en cada categoría de la tercera macrocategoría.....	373
Ilustración 101:	Relación del motor RBMT1 en cada categoría de la tercera macrocategoría.....	374
Ilustración 102:	Relación del motor RBMT2 en cada categoría de la tercera macrocategoría.....	375
Ilustración 103:	Relación tipo de motor categorías de la cuarta macrocategoría sin carga lingüística	376
Ilustración 104:	Relación del motor SMT1 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica.....	377
Ilustración 105:	Relación del motor SMT2 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica.....	379
Ilustración 106:	Relación del motor RBMT1 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica	380
Ilustración 107:	Relación del motor RBMT2 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica	381
Ilustración 108:	Relación tipo de motor categorías de la cuarta macrocategoría con carga semántica	382
Ilustración 109:	Relación del motor SMT1 en cada categoría de la cuarta macrocategoría en las palabras sin carga semántica.....	383
Ilustración 110:	Relación del motor SMT2 en cada categoría de la cuarta macrocategoría en las palabras con carga semántica	385

Índice de tablas

Ilustración 111:	Relación del motor RBMT1 en cada categoría de la cuarta macrocategoría en palabras sin carga semántica.....	386
Ilustración 112:	Relación del motor RBMT2 en cada categoría de la cuarta macrocategoría en palabras sin carga semántica.....	387
Ilustración 113:	Relación tipo de motor categorías de la quinta macrocategoría	388
Ilustración 114:	Relación de los valores de errores sintácticos en la traducción de compuestos nominales.....	398
Ilustración 115:	Relación de los valores de errores léxicos en la traducción de compuestos nominales.....	399
Ilustración 116:	Relación de los valores la traducción de verbos alemanes separables	402

Índice de tablas

Tabla 1:	Niveles del lenguaje especializado según Hoffmann.....	27
Tabla 2:	Categorías de la traducción técnica según Gamero	28
Tabla 3:	Clasificación de la traducción técnica escrita de Gamero (2001, 69).....	30
Tabla 4:	Factores externos que pueden generar problemas pragmáticos según Nord (2001a, 151-68).....	32
Tabla 5:	Clasificación de los problemas de traducción según Nord	35
Tabla 6:	Tipos de problemas según Krings.....	36
Tabla 7:	Clasificación de los problemas de Krings según el nivel en que se presentan	36
Tabla 8:	Clasificación de los problemas de la traducción según Schmitt	38
Tabla 9:	Clasificación de los problemas de traducción según Hurtado	40
Tabla 10:	Tipología de errores de Nord.....	41
Tabla 11:	Clasificación de los errores e inadecuaciones de traducción según Hurtado.....	43
Tabla 12:	Clasificación de errores de Schmitt	47
Tabla 13:	Ejemplo del uso de oraciones relativas en castellano y alternativas en alemán.....	53
Tabla 14:	Convenciones formales en alemán (Nord 2001a, 189).....	53
Tabla 15:	Ejemplo de ambigüedad de Uszkoreit	55
Tabla 16:	Ejemplos de homonimia total en alemán	56
Tabla 17:	Ejemplo de homónimos parciales en alemán (Löbner 2003, 58-59)	56
Tabla 18:	Ambigüedad semántica en una oración (Ernst 2003: 79)	57
Tabla 19:	Ambigüedad de nombres compuestos en alemán	57
Tabla 20:	Ejemplo de polisemia en alemán	58
Tabla 21:	Tipos de polisemiaMetáforas y metonimias	59
Tabla 22:	Metáforas en alemán.....	60
Tabla 23:	Ejemplos de ambigüedad en sustantivos compuestos en alemán de Hutchins y Somers ..	61
Tabla 24:	Ambigüedad léxica debido a la estructura morfológica en alemán	62
Tabla 25:	Ejemplos de ambigüedad sintáctica.....	62
Tabla 26:	Ejemplo de una oración ambigua Garden Path (Bader 2004, 8-9)	63

Índice de tablas

Tabla 27:	Ejemplos de ambigüedad nominalización + referencia + sustantivo	63
Tabla 28:	Construcciones ambiguas con sustantivo en plural y verbo en infinitivo	63
Tabla 29:	Ambigüedad estructural	64
Tabla 30:	<i>Konjunktiv I</i> en alemán	65
Tabla 31:	<i>Konjunktiv II</i>	65
Tabla 32:	Verbo ser / estar	66
Tabla 33:	Formas verbales compuestas y posición del verbo	67
Tabla 34:	Ejemplo de la diferencia en el ordenamiento gramatical	67
Tabla 35:	Ejemplo de la diferencia en el ordenamiento gramatical	67
Tabla 36:	Diferencias en la colocación de los complementos en alemán y castellano	68
Tabla 37:	Diferencias en la colocación de la negación en alemán y castellano	68
Tabla 38:	Ejemplo de nombres compuestos en alemán con diferentes significados.....	69
Tabla 39:	Verbos compuestos en alemán con diferente significado	70
Tabla 40:	Uso de las declinaciones en alemán.....	70
Tabla 41:	Declinación del artículo indefinido en alemán	71
Tabla 42:	Nominalización con <i>-ung</i> y preposiciones	72
Tabla 43:	Tipos de problemas que se presentan en el análisis según Arnold	100
Tabla 44:	Problemas de transferencia según Arnold.....	101
Tabla 45:	Ejemplos de problemas en la transferencia.....	101
Tabla 46:	Ejemplos de problemas de orden morfológico	103
Tabla 47:	Diferencias en el número y género	103
Tabla 48:	Ejemplos de problemas en la traducción de sustantivos compuestos (Hutchins y Somers 1992, 85).....	104
Tabla 49:	Problemas en la traducción automática de homógrafos.....	105
Tabla 50:	Ejemplo de diferencia léxica entre el inglés, francés y castellano.....	107
Tabla 51:	Problemas léxicos de orden terminológico	108
Tabla 52:	Problemas léxicos en la traducción automática de preposiciones (Ramlow 2009, 185)	108
Tabla 53:	Problemas léxicos en la traducción automática de <i>phrasen</i>	109
Tabla 54:	Ejemplo de Ramlow de interpretaciones posibles en una frase idiomática	109
Tabla 55:	Problemas que pueden aparecer en la traducción automática de frases idiomáticas	110
Tabla 56:	Expresiones idiomáticas correspondientes en alemán y castellano	110
Tabla 57:	Anáforas.....	111
Tabla 58:	Diferencia de tratamiento de anáforas en inglés y francés.....	112
Tabla 59:	Ambigüedad por ampliación de cuantificadores.....	113
Tabla 60:	Problemas por la omisión o falta de traducción de palabras.....	113
Tabla 61:	Otros problemas de orden léxico	114
Tabla 62:	Problemas en el ordenamiento de las palabras	114

Índice de tablas

Tabla 63:	Problemas en la ordenación de oraciones relativas.....	115
Tabla 64:	Problemas debidos a las diferencias en la estructura sintáctica.....	115
Tabla 65:	Problemas sintácticos en la traducción de preposiciones.....	116
Tabla 66:	Ejemplo del uso diferente de las preposiciones en castellano e inglés.....	116
Tabla 67:	Problemas en la referencia sintáctica.....	116
Tabla 68:	Ejemplo de ambigüedad estructural.....	117
Tabla 69:	Ambigüedad sintáctica en determinadas construcciones.....	117
Tabla 70:	Tipos de divergencias según Dorr.....	118
Tabla 71:	Ejemplo de diferencias estructurales.....	119
Tabla 72:	Otro tipo de problemas de orden sintáctico.....	119
Tabla 73:	Problemas de pronominalización.....	120
Tabla 74:	Errores de puntuación.....	120
Tabla 75:	Ejemplo de ordenamiento sintáctico incorrecto.....	123
Tabla 76:	Ejemplo de errores en la conjugación de los verbos en los motores SMT.....	123
Tabla 77:	Ejemplo de error en la identificación de categorías.....	126
Tabla 78:	Ejemplo de error en la conjugación de los verbos.....	126
Tabla 79:	Ejemplo de error en la traducción de una expresión idiomática.....	126
Tabla 80:	Clasificación de errores de Green.....	128
Tabla 81:	Clasificación de errores de Lavorel.....	129
Tabla 82:	Clasificación de errores de Löffler-Laurian.....	131
Tabla 83:	Categorías de errores de LISA QA v3.1.....	132
Tabla 84:	Las categorías de errores más importantes en LISA.....	132
Tabla 85:	Clasificación de errores de SAE J2450.....	134
Tabla 86:	Clasificación de errores del estudio realizado por O'Brien.....	139
Tabla 87:	Ejemplo para el cálculo de BLEU (Sánchez-Martínez 2014).....	142
Tabla 88:	Ejemplo para el cálculo de WER (Sánchez-Martínez 2014).....	145
Tabla 89:	Ejemplo para el cálculo de WER (Sánchez-Martínez 2014).....	146
Tabla 90:	Categorías de errores que tienen que ver con <i>Accuracy</i>	156
Tabla 91:	Categorías de errores que tienen que ver con <i>Design</i>	158
Tabla 92:	Tipos de errores que tienen que ver con <i>Fluency</i>	161
Tabla 93:	Tipos de errores de la categoría <i>Verity</i>	162
Tabla 94:	Tipos de errores de la categoría <i>Terminology</i>	164
Tabla 95:	Tipos de errores de la categoría <i>Style</i>	165
Tabla 96:	Tipos de errores de la categoría <i>Locale convention</i>	166
Tabla 97:	Tipos de errores de la categoría <i>Internationalization</i>	169
Tabla 98:	Ejemplo de RPE (Schumann y Sivaloganathan 2013).....	176
Tabla 99:	Ejemplo de MPE (Schumann y Sivaloganathan 2013).....	177

Índice de tablas

Tabla 100:	Ejemplo de CPE Traducción automática	177
Tabla 101:	Ejemplo comparativo de los tipos principales de posesición	177
Tabla 102:	Diferencias entra la posesición rápida y completa según TAUS.....	178
Tabla 103:	Herramientas para la posesición (Flórez 2012, 221)	188
Tabla 104:	Comparación de las definiciones tradicionales de los problemas de traducción	191
Tabla 105:	Parámetros de MQM.....	212
Tabla 106:	Fases del proyecto TaraXÚ.....	216
Tabla 107:	Etiquetas encontradas en el archivo translations.xml	217
Tabla 108:	Etiquetas encontradas en el archivo posediting-results.xml	218
Tabla 109:	Lista de subcorpus encontrados	219
Tabla 110:	Elementos empleados para la organización del corpus.....	221
Tabla 111:	Descripción de la tabla para las anotaciones en la primera fase	223
Tabla 112:	Errores marcados en las macrocategorías de MQM	230
Tabla 113:	Errores y puntos problemáticos más frecuentes hallados en los originales	237
Tabla 114:	Errores generales más frecuentes encontrados en las traducciones	241
Tabla 115:	Errores de carácter formal (no lingüístico)	243
Tabla 116:	Omisiones	244
Tabla 117:	Adiciones	246
Tabla 118:	Errores en la traducción de sustantivos.....	249
Tabla 119:	Tipos de errores en la traducción de adjetivos.....	250
Tabla 120:	Tipos de errores en la traducción de los verbos	252
Tabla 121:	Errores en la traducción de otros tipos de palabras.....	253
Tabla 122:	Errores más frecuentes detectados en el primer análisis.....	255
Tabla 123:	Ejemplos de los errores encontrados en la SMT1	257
Tabla 124:	Ejemplos de los errores encontrados en las propuestas de SMT2	259
Tabla 125:	Ejemplos de los errores encontrados en las propuestas de RBMT1	261
Tabla 126:	Ejemplos de errores encontrados en las propuestas de RBMT2.....	263
Tabla 127:	Comparación de los errores más frecuentes y los motores	264
Tabla 128:	Comparación de los errores más frecuentes y los motores empleados	265
Tabla 129:	Ejemplos en los que el poseedor prefirió traducir que poseer.....	266
Tabla 130:	Ejemplo de selección de una propuesta en particular teniendo otras opciones iguales ...	267
Tabla 131:	Ejemplo de selección de una propuesta con una distancia mayor según Levenshtein.....	267
Tabla 132:	Número de errores encontrados en la categoría <i>Accuracy</i>	269
Tabla 133:	Ejemplos de errores encontrados en la categoría <i>Accuracy</i>	270
Tabla 134:	Ejemplos de los errores encontrados en la subcategoría <i>Mistranslation</i>	272
Tabla 135:	Número de errores encontrados en la categoría <i>Design</i>	273
Tabla 136:	Ejemplos de errores encontrados en la categoría <i>Design</i>	273

Índice de tablas

Tabla 137: Número de errores encontrados en la categoría <i>Fluency</i>	274
Tabla 138: Errores de la subcategoría <i>Grammar</i>	276
Tabla 139: Errores de las subcategorías <i>Spelling</i> y <i>Typography</i>	277
Tabla 140: Errores de la subcategoría <i>Unintelligible</i>	277
Tabla 141: Errores de la subcategoría <i>Locale covention</i>	278
Tabla 142: Errores de la subcategoría <i>Style</i>	279
Tabla 143: Errores de la subcategoría <i>Terminology</i>	280
Tabla 144: Errores de la subcategoría <i>Verity</i>	281
Tabla 145: Descripción de las columnas de la hoja de cálculo que utilizamos para la segunda fase del análisis	284
Tabla 146: Macrocategorías y categorías de errores empleadas en la segunda fases del análisis	287
Tabla 147: Descripción de los tipos de errores	288
Tabla 148: Definición de la categoría presencia de léxico extranjero	289
Tabla 149: Definición de la categoría tratamiento de nombres propios	290
Tabla 150: Definición de la subcategoría de errores en la conversión de convenciones	290
Tabla 151: Definición de la subcategoría errores de funcionalidad de elementos tipográficos	291
Tabla 152: Definición de la subcategoría traducción errónea de enlaces, hipertexto y referencias	291
Tabla 153: Definición de la subcategoría traducción de verbos compuestos separables	293
Tabla 154: Definición de la subcategoría traducción de conjunciones y partículas	293
Tabla 155: Definición de la subcategoría error de transferencia sintáctica en composiciones	294
Tabla 156: Definición de la subcategoría errores en la traducción de oraciones subordinadas	295
Tabla 157: Definición de la subcategoría errores en la traducción de oraciones coordinadas	295
Tabla 158: Definición de la subcategoría errores en la traducción de negaciones	296
Tabla 159: Definición de la subcategoría errores en la traducción de colocaciones o fraseologías	296
Tabla 160: Definición de la subcategoría errores en la flexión de género	297
Tabla 161: Definición de la subcategoría errores en la flexión de número	297
Tabla 162: Definición de la subcategoría errores en la flexión verbal (tiempo)	298
Tabla 163: Definición de la subcategoría errores en la flexión verbal (modo)	298
Tabla 164: Definición de la subcategoría errores en la flexión verbal (persona)	299
Tabla 165: Definición de la subcategoría errores en la flexión verbal (voz)	299
Tabla 166: Definición de la subcategoría errores de estilo	300
Tabla 167: Definición de la subcategoría errores ortográficos	300
Tabla 168: Definición de la subcategoría errores tipográficos	300
Tabla 169: Definición de la subcategoría errores de sintaxis	301
Tabla 170: Definición de la subcategoría errores en la traducción de preposiciones	302
Tabla 171: Definición de la subcategoría errores en la traducción de sintagmas preposicionales	302
Tabla 172: Definición de la subcategoría errores en la traducción de determinantes	303

Índice de tablas

Tabla 173:	Definición de la subcategoría errores en la traducción de adverbios.....	303
Tabla 174:	Definición de la subcategoría errores en la traducción de estructuras adverbiales.....	303
Tabla 175:	Definición de la subcategoría errores en la traducción de conjunciones y partículas.....	304
Tabla 176:	Definición de la subcategoría errores en la traducción de sustantivos	304
Tabla 177:	Definición de la subcategoría errores en la traducción de adjetivos.....	305
Tabla 178:	Definición de la subcategoría errores en la traducción de verbos.....	305
Tabla 179:	Error en la traducción de formas verbales compuestas	306
Tabla 180:	Definición de la subcategoría error de transferencia sintáctica en compuestos nominales	307
Tabla 181:	Definición de la subcategoría error de transferencia sintáctica léxica en los compuestos nominales	307
Tabla 182:	Error en la transferencia léxica de compuestos nominales y composiciones.....	308
Tabla 183:	Definición de la subcategoría errores causados por la presencia de etiquetas en el original	309
Tabla 184:	Definición de la subcategoría errores causados por la redacción del original	309
Tabla 185:	Definición de la subcategoría errores causados por errores orto-tipográficos en el original	310
Tabla 186:	Otro tipo de errores	310
Tabla 187:	Segmento que se tradujo de cero	315
Tabla 188:	Ejemplos de errores más frecuentes independientemente del motor seleccionado.....	317
Tabla 189:	Ejemplos de errores más frecuentes en las propuestas del smt1	319
Tabla 190:	Ejemplos de errores más frecuentes en las propuestas del smt2.....	322
Tabla 191:	Ejemplos de errores más frecuentes en las propuestas del rbmt1	325
Tabla 192:	Ejemplos de errores más frecuentes en las propuestas del rbmt2	327
Tabla 193:	Segmentos en que encontramos errores sin justificación.....	330
Tabla 194:	Errores detectados en la posesición.....	333
Tabla 195:	descripción general de las macrocategorías empleadas en el análisis.....	339
Tabla 196:	Número y porcentaje de errores marcados en la primera macrocategoría	341
Tabla 197:	Número y porcentaje de errores marcados en la segunda macrocategoría	342
Tabla 198:	Número y porcentaje total de errores marcados en la tercera macrocategoría	343
Tabla 199:	Porcentaje total de errores marcados en la cuarta macrocategoría en palabras sin carga semántica	345
Tabla 200:	Porcentaje total de errores marcados en la cuarta macrocategoría en palabras con carga semántica	346
Tabla 201:	Número y porcentaje total de errores marcados en la quinta categoría	348
Tabla 202:	Número y porcentaje total de errores marcados en la sexta categoría	349
Tabla 203:	Número total de segmentos seleccionados por macrocategoría por tipo de sistema	351
Tabla 204:	Número y porcentaje total de errores marcados en la primera macrocategoría en el motor SMT1.....	361

Índice de tablas

Tabla 205: Número y porcentaje total de errores marcados en la primera macrocategoría en el motor SMT2.....	362
Tabla 206: Número y porcentaje total de errores marcados en la primera macrocategoría en el motor RBMT1	364
Tabla 207: Número y porcentaje total de errores marcados en la primera macrocategoría en el motor RBMT2	365
Tabla 208: Categorías con más número de ediciones por motor	366
Tabla 209: Número y porcentaje total de errores marcados en la segunda macrocategoría en el motor SMT1.....	367
Tabla 210: Número y porcentaje total de errores marcados en la segunda macrocategoría en el motor SMT2.....	368
Tabla 211: Número y porcentaje total de errores marcados en la segunda macrocategoría en el motor RBMT1	369
Tabla 212: Número y porcentaje total de errores marcados en la segunda macrocategoría en el motor RBMT2	370
Tabla 213: Categorías con más número de ediciones por motor	370
Tabla 214: Número y porcentaje total de errores marcados en la tercera macrocategoría en el motor SMT1.....	372
Tabla 215: Número y porcentaje total de errores marcados en la tercera macrocategoría en el motor SMT2.....	373
Tabla 216: Número y porcentaje total de errores marcados en la tercera macrocategoría en el motor RBMT1	374
Tabla 217: Número y porcentaje total de errores marcados en la tercera macrocategoría en el motor RBMT2	375
Tabla 218: Categorías con más número de ediciones por motor	376
Tabla 219: Número y porcentaje total de errores en la cuarta macrocategoría (sin carga semántica) en el motor SMT1	377
Tabla 220: Número y porcentaje total de errores en la cuarta categoría de palabras sin carga semántica en el motor SMT2.....	378
Tabla 221: Número y porcentaje total de errores en la cuarta categoría de palabras sin carga semántica en el motor RBMT1	379
Tabla 222: Número y porcentaje total de errores en la cuarta categoría de palabras sin carga semántica en el motor RBMT2.....	380
Tabla 223: Categorías con más número de ediciones por motor	381
Tabla 224: Número y porcentaje total de errores en la cuarta macrocategoría (con carga semántica) en el motor SMT1.....	383
Tabla 225: Número y porcentaje total de errores en la cuarta categoría de palabras con carga semántica en el motor SMT2.....	384
Tabla 226: Número y porcentaje total de errores en la cuarta categoría de palabras con carga semántica en el motor SMT2.....	385
Tabla 227: Número y porcentaje total de errores en la cuarta categoría de palabras con carga semántica en el motor RBMT2.....	386

Índice de tablas

Tabla 228: Categorías con más número de ediciones por motor	388
Tabla 229: Número y porcentaje total de errores en la quinta categoría en el motor SMT1	389
Tabla 230: Número y porcentaje total de errores en la quinta categoría en el motor SMT2	389
Tabla 231: Número y porcentaje total de errores en la quinta categoría en el motor RBMT1	389
Tabla 232: Número y porcentaje total de errores en la quinta categoría en el motor RBMT2	389
Tabla 233: Categorías con más número de ediciones por motor	390
Tabla 234: Número y porcentaje de los valores asignados a las ediciones de los verbos.....	391
Tabla 235: Ejemplo de edición de errores en la traducción de verbos.....	392
Tabla 236: Valores de los errores en la traducción de sustantivos.....	393
Tabla 237: Propuestas similares rechazadas en comparación con las seleccionadas para la posedición	395
Tabla 238: En la PE se evita la edición de etiquetas.....	396
Tabla 239: En la PE se evita la traducción de <i>untranslated</i>	396
Tabla 240: En la PE prefiere realizar ediciones en el orden sintáctico	397
Tabla 241: Valores de los errores sintácticos en la traducción de compuestos nominales	398
Tabla 242: Valores de los errores en la transferencia léxica de compuestos nominales.....	398
Tabla 243: Tratamiento de las etiquetas y edición de compuestos nominales.....	400
Tabla 244: Propuestas con el mismo valor LV de dos motores diferentes	400
Tabla 245: En la PE se seleccionan propuestas sin adiciones u omisiones.....	401
Tabla 246: Puntuación correcta.....	401
Tabla 247: Valores de los errores en la transferencia de verbos compuestos separables	402
Tabla 248: Ejemplo de las ediciones en errores de verbos separables en dos propuestas diferentes.	403
Tabla 249: Ejemplo de las ediciones en errores de verbos sepables en dos propuestas diferentes....	404
Tabla 250: Ejemplo de propuestas similares, con puntuación diferente	404
Tabla 251: Ejemplo de propuestas similares, con problemas de concordancia de número	405

PARTE V: ANEXOS

Los siguientes anexos se encuentran en forma digital en el CD-Rom que acompaña la tesis:

Anexo 1: Acuerdo de derechos con la firma Euroscript GmbH, Alemania

Anexo 2: Datos originales del corpus (ver archivos translation.xml y posediting.xml)

Anexo 3: Tabla de análisis y conteo de errores primera fase

Anexo 4: Tabla de análisis con Levenshtein

Anexo 5: Tabla de análisis y conteo de errores primera fase

Anexo 6: Tabla de análisis y conteo de errores con las categorías de MQM

Anexo 7: Definición de las categorías de errores de MQM