

Tesis Doctoral



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

Evaluación de la calidad de la  
traducción automática de reseñas  
turísticas en línea desde la  
perspectiva de la localización

**Autora:** Maria Carmen Rosa-Sorlózano

**Director:** Miguel Ángel Candel-Mora

**DEPARTAMENTO DE LINGÜÍSTICA APLICADA**

**Valencia, marzo 2024**

## Agradecimientos

Agradezco sinceramente a todas las personas que me han acompañado durante la elaboración de esta tesis y todo lo que conlleva el doctorado. Aunque me siento muy orgullosa de mí misma por haber llegado hasta aquí, sé que sin su apoyo incondicional no habría sido posible.

En primer lugar, me gustaría expresar mi gratitud a mi familia por su apoyo y amor incondicional a lo largo de este arduo, aunque gratificante proceso.

También me gustaría agradecer a mi tutor, Miguel Ángel Candel Mora, su dedicación, su esfuerzo y su orientación, los cuales han sido esenciales para completar esta tesis y para mi desarrollo académico.

A mis queridas amigas y a mi pareja, que han aguantado interminables charlas sobre el progreso de mi doctorado y mis preocupaciones, a veces a horas intempestivas. Solo puedo estar agradecida por su apoyo inquebrantable, sus palabras de aliento y su compañía incondicional. De igual forma, no puedo dejar de agradecer a mi fiel compañera canina, Lluna, cuya presencia ha sido una fuente constante de alegría durante los momentos más solitarios.

Por supuesto, estoy también profundamente agradecida con todas aquellas personas que, de una forma u otra, han contribuido a la realización de esta tesis doctoral. Finalmente, me dedico este logro a mí misma, por aguantar y creer que podía, incluso cuando dudaba de mis capacidades. Ha sido indudablemente un viaje duro, pero muy satisfactorio que repetiría de nuevo sin ápice de duda. No puedo esperar a ver qué depara mi futuro profesional y académico.

Por todo lo expuesto, gracias, de todo corazón.

## Resumen

Los usuarios de internet han pasado a ser contribuidores activos en la Web 2.0. Estudios recientes revelan que siete de cada 10 usuarios de internet en todo el mundo confían en la opinión y las reseñas publicadas en línea por otros usuarios. De la misma forma, según estadísticas recientes de las agencias españolas de turismo, el uso de internet ha crecido más de un 29 %: casi todos los usuarios (99,2 %) lo usan para buscar información, un 76,5 % lo usa para hacer reservas y el 52,4 % para el pago de servicios. A pesar del gran potencial y el volumen de negocio de este sector, las plataformas de reseñas (tanto de viajes como de reservas de restaurantes) solo utilizan la traducción automática (TA) de las opiniones que los consumidores dejan en dichas páginas sin procesamiento ni revisión. En paralelo, los estudios de traducción han concedido un papel clave durante los últimos 30 años a la subdisciplina de la localización, consistente en adaptar el mensaje a las preferencias lingüísticas y culturales específicas del usuario. Por este motivo, esta investigación se centra en el análisis de un corpus de reseñas traducidas automáticamente para identificar patrones de error y sus efectos en la calidad del texto según parámetros estudiados en la localización. Áreas como el turismo, las finanzas y el marketing podrían verse beneficiadas si mejoran sus procesos de traducción, puesto que un mejor entendimiento de los servicios anunciados facilita la interacción del consumidor con dichos productos y servicios.

## Resum

Els usuaris d'internet han passat a ser contribuïdors actius en la Web 2.0. Estudis recents revelen que set de cada 10 usuaris d'internet a tot el món confien en l'opinió i les ressenyes publicades en línia per altres usuaris. De la mateixa forma, segons estadístiques recents de les agències espanyoles de turisme, l'ús d'internet ha crescut més d'un 29%: quasi tots els usuaris (99,2%) l'utilitzen per a buscar informació, un 76,5% l'usa per a fer reserves i el 52,4% per al pagament de servicis. Malgrat el gran potencial i el volum de negoci d'este sector, les plataformes de ressenyes (tant de viatges com de reserves de restaurants) només utilitzen la traducció automàtica (TA) de les opinions que els consumidors deixen en estes pàgines sense processament ni revisió. En paral·lel, els estudis de traducció han concedit un paper clau durant els últims 30 anys a la subdisciplina de la localització, consistent a adaptar el missatge a les preferències lingüístiques i culturals específiques de l'usuari. Per aquest motiu, esta investigació se centra en l'anàlisi d'un corpus de ressenyes traduïdes automàticament per a identificar patrons d'error i els seus efectes en la qualitat del text segons paràmetres estudiats en la localització. Àrees com el turisme, les finances i el màrqueting podrien veure's beneficiades si milloren els seus processos de traducció, ja que un millor enteniment dels servicis anunciats facilita la interacció del consumidor amb aquests productes i servicis.

## Abstract

Internet users have become active contributors to the Web 2.0. Recent studies reveal that seven out of 10 internet users worldwide rely on the opinions and reviews posted online by other users. Similarly, according to recent statistics from the main Spanish tourism agencies, Internet usage has grown by more than 29%: almost all users (99.2%) use it to search for information, 76.5% use it to make reservations and 52.4% use it to pay for services. Despite the great potential and turnover of this sector, review platforms (both for travel and restaurant reservations) only use machine translation (MT) of the reviews that consumers leave on these pages without processing or review. In parallel, translation studies have attributed a major role over the last 30 years to the sub-discipline of localization, which consists of adapting the message to the user's specific linguistic and cultural preferences. For this reason, this research focuses on the analysis of a corpus of machine-translated reviews to identify error patterns and their effects on text quality according to parameters studied in localization. Areas such as tourism, finance and marketing could benefit from improving their translation processes, since a better understanding of the advertised services facilitates consumer interaction with those services.

## Tabla de contenido

Índice de tablas.....	7
Índice de ilustraciones.....	10
Índice de ecuaciones .....	11
<b>1. Introducción.....</b>	<b>12</b>
1.1. Premisas de trabajo .....	13
1.2. Objetivos.....	14
1.3. Estructura del trabajo .....	14
<b>2. Marco Teórico.....</b>	<b>21</b>
2.1. Contenido Generado por el Usuario (CGU).....	21
2.1.1. Antecedentes.....	21
2.1.2. Definición y características de CGU .....	28
2.1.3. Importancia y estudio del CGU .....	31
2.1.4. Importancia del CGU .....	32
2.1.5. Tipos de CGU: las RGU como nuevo género textual.....	44
2.2. La Localización .....	49
2.2.1. Orígenes.....	49
2.2.2. Definición.....	51
2.2.3. Características de la localización.....	55
2.2.4. Traducción vs. Localización .....	56
2.3. Traducción Automática .....	59
2.3.1. Definición.....	62
2.3.2. Tipos de TA .....	67
2.3.3. La posesición .....	101
2.3.4. Métricas de calidad de TAN.....	103
2.3.5. Orígenes y evolución de la TA .....	104
2.3.6. Reflexiones finales sobre TA .....	123
<b>3. Análisis y resultados.....</b>	<b>125</b>
3.1. Metodología .....	125
3.1.1. Corpus de reseñas bilingüe (Access) .....	127
3.1.2. Corpus monolingüe (Excel).....	129
3.1.3. Listas de frecuencias de WordSmith .....	132
3.1.4. Tablas para el análisis de la longitud de las reseñas .....	137
3.2. Análisis cuantitativo y cualitativo del corpus de reseñas sobre restaurantes traducidas con TA .....	139
3.2.1. Análisis cuantitativo de aspectos de localización en el corpus de reseñas de restaurantes.....	139
3.2.2. Análisis cualitativo de los resultados de la TA en el corpus de reseñas de restaurantes	142
3.3. Análisis cuantitativo y cualitativo del corpus de reseñas sobre hoteles traducidas con TA	154
3.3.1. Análisis cuantitativo de aspectos de localización en el corpus de reseñas de hoteles	155
3.3.2. Análisis cualitativo de los resultados de la TA en el corpus de reseñas de hoteles ....	159

3.4.	Comparativa de los resultados de los corpus bilingües de reseñas de restaurantes y de hoteles.....	178
3.5.	Comparativa de la naturalidad y autenticidad en las reseñas escritas originalmente en español.....	184
3.5.1.	Comparativa sobre la naturalidad y la autenticidad en el corpus de reseñas sobre restaurantes. Corpus monolingüe y bilingüe .....	185
3.5.2.	Comparativa sobre la naturalidad y la autenticidad en el corpus de reseñas sobre hoteles. Corpus monolingüe y bilingüe .....	193
3.6.	Observaciones análisis cuantitativo y cualitativo .....	201
3.7.	Análisis de relación entre la longitud de las reseñas y la calidad del resultado de la TA	204
3.7.1.	Parámetros de calidad utilizados .....	206
3.7.2.	Calidad de las reseñas de hoteles según su longitud .....	208
3.7.3.	Calidad de las reseñas de restaurantes según su longitud.....	210
3.8.	Observaciones análisis de la longitud de la reseña y calidad .....	212
4.	<i>Conclusiones</i> .....	214
4.1.	Trabajo futuro .....	218
5.	<i>Bibliografía</i> .....	220

## Índice de tablas

Tabla 1. Ejemplos de clasificación de las reseñas (Elaboración propia).....	16
Tabla 2. Porcentaje de usuarios de Internet en los últimos 3 meses por tipo de actividad realizada. 2022 (% de personas de 16 a 74 años. ....	26
Tabla 3. Fases de la experiencia temporal de un cliente (Vásquez, 2012: 109). ....	41
Tabla 4. Tabla de siglas comunes en TA (Elaboración propia). ....	61
Tabla 5. Ventajas e inconvenientes de la TA basada en reglas (Elaboración propia). ....	68
Tabla 6. Principales diferencias entre la TAE y la TA basada en reglas (Elaboración propia). ....	76
Tabla 7. Corpus monolingüe (Excel). ....	129
Tabla 8. Lista de frecuencias de palabras encontradas en las reseñas de hoteles (Corpus ES-TA). ....	132
Tabla 9. Lista de frecuencias de palabras encontradas en las reseñas de hoteles (Corpus ES Original).....	133
Tabla 10. Lista de frecuencias de palabras encontradas en las reseñas de restaurantes (ES-TA). ....	134
Tabla 11. Lista de frecuencias de palabras encontradas en las reseñas de hoteles (Corpus ES Original).....	135
Tabla 12. Tabla sobre la longitud de las reseñas de hoteles en todos los corpus. ....	137
Tabla 13. Tabla sobre la longitud de las reseñas de restaurantes en todos los corpus. ....	138
Tabla 14. Datos sobre número de reseñas sobre restaurantes y los principales tipos de errores de localización. ....	139
Tabla 15. Errores con la traducción del verbo to have por «tener». ....	140
Tabla 16. Errores con la traducción del verbo to be por «ser/estar». ....	141
Tabla 17. Errores en la traducción de expresiones idiomáticas.....	141
Tabla 18. Errores de localización de tipo cultural. ....	142
Tabla 19. Clasificación de la calidad de la TA en función del número y porcentaje de errores lingüísticos. Corpus restaurantes. ....	144
Tabla 20. Diferentes categorías de errores lingüísticos. ....	145
Tabla 21. Clasificación de la calidad de la TA en función del número y porcentaje de errores culturales. Corpus restaurantes. ....	147
Tabla 22. Errores culturales. Ejemplos de una misma palabra en inglés traducida inconsistentemente. Corpus restaurantes.....	148
Tabla 23. Errores culturales. Ejemplos de errores con los nombres de platos regionales. Corpus restaurantes. ....	148
Tabla 24. Número problemas que afectan al nivel de calidad de la TA. Corpus restaurantes. ....	150
Tabla 25. Ejemplos de errores de duplicidad o explicación encontrados. Corpus de restaurantes. ....	151
Tabla 26. Ejemplos de errores causados por la calidad del motor de TA. Corpus de restaurantes. ....	152
Tabla 27. Ejemplos de errores causados por errores en la reseña original en inglés. Corpus de restaurantes. ....	152
Tabla 28. Ejemplos de errores por palabras mal escritas. Corpus de restaurantes.....	154
Tabla 29. Datos sobre número de reseñas sobre hoteles y los principales tipos de errores de localización. ....	155
Tabla 30. Ejemplos de lugares localizados incorrectamente. ....	156
Tabla 31. Traducción incorrecta del verbo to have. Corpus hoteles. ....	156
Tabla 32. Traducción incorrecta del verbo to be. Corpus hoteles. ....	157



Tabla 33. Clasificación de la calidad de la TA en función del número y porcentaje de errores lingüísticos. Corpus hoteles.....	160
Tabla 34. Ejemplos de errores de localización lingüísticos. Corpus de hoteles. ....	161
Tabla 35. Traducción incorrecta de frases hechas o expresiones en inglés. Corpus de hoteles. ....	164
Tabla 36. Traducción incorrecta de oraciones pasivas y de las oraciones de relativo. Corpus de hoteles.....	165
Tabla 37. Clasificación de la calidad de la TA en función del número y porcentaje de errores culturales. Corpus hoteles.....	167
Tabla 38. Errores de localización culturales al traducir nombres de lugares turísticos de la ciudad.....	168
Tabla 39. Errores de localización cultural por traducción incorrecta de ingredientes o platos típicos. Corpus hoteles. ....	169
Tabla 40. Errores de localización culturales por la no traducción de expresiones en inglés o la traducción incorrecta de festividades locales. Corpus de hoteles.....	169
Tabla 41. Errores de localización culturales por traducción incorrecta de abreviaturas, signos o divisas. Corpus de hoteles. ....	171
Tabla 42. Errores culturales de localización por uso de variantes no propias del español de España. Corpus de hoteles. ....	171
Tabla 43. Número problemas que afectan al nivel de calidad de la TA. Corpus hoteles. ....	172
Tabla 44. Ejemplos de errores de duplicidad o explicación encontrados. Corpus de hoteles..	173
Tabla 45. Traducción incorrecta de los verbos to be y to have. Corpus hoteles. ....	174
Tabla 46. Ejemplos de errores causados por la calidad del motor de TA. Corpus hoteles. ....	176
Tabla 47. Ejemplos de errores causados por errores en la reseña original en inglés. Corpus hoteles.....	177
Tabla 48. Casos de problemas culturales o lingüísticos causados por diferentes tipos de errores. Corpus restaurantes.....	179
Tabla 49. Casos de problemas culturales o lingüísticos causados por diferentes tipos de errores. Corpus de hoteles. ....	179
Tabla 50. Errores causados por la traducción incorrecta de oraciones de relativo.....	180
Tabla 51. Errores por detección incorrecta de las mayúsculas.....	180
Tabla 52. Errores causados por la emisión de los emoticonos en la TA de las reseñas.....	182
Tabla 53. Errores causados por un uso abusivo de exclamaciones. ....	182
Tabla 54. Palabras más frecuentes en reseñas de restaurantes traducidas con TA al ES.....	185
Tabla 55. Adjetivos más frecuentes en reseñas de restaurantes traducidas con TA al ES. ....	187
Tabla 56. Palabras más frecuentes en reseñas de restaurantes corpus monolingüe.....	189
Tabla 57. Adjetivos más frecuentes en reseñas de restaurantes corpus monolingüe. ....	191
Tabla 58. Palabras con frecuencia similar en ambos corpus. Corpus restaurantes.....	192
Tabla 59. Adjetivos con frecuencia similar en ambos corpus. Corpus restaurantes. ....	193
Tabla 60. Palabras más frecuentes en reseñas de hoteles traducidas con TA al ES. ....	194
Tabla 61. Adjetivos más frecuentes en reseñas de hoteles traducidas con TA al ES.....	195
Tabla 62. Palabras más frecuentes en reseñas de hoteles corpus monolingüe. ....	197
Tabla 63. Adjetivos más frecuentes en reseñas de hoteles corpus monolingüe. ....	199
Tabla 64. Palabras con frecuencia similar en ambos corpus. Corpus hoteles. ....	199
Tabla 65. Adjetivos con frecuencia similar en ambos corpus. Corpus hoteles. ....	201
Tabla 66. Datos generales sobre reseñas.....	205
Tabla 67. Ejemplos de reseñas y longitudes. ....	205
Tabla 68. Ejemplos de errores por omisiones o uso incorrecto de signos de puntuación. ....	207

Tabla 69. Porcentajes de reseñas según su calidad y longitud. Corpus hoteles. ....	209
Tabla 70. Porcentajes de reseñas según su calidad y longitud. Corpus restaurantes. ....	211
Tabla 71. Porcentajes de reseñas inaceptables de restaurantes vs. Hoteles. ....	212
Tabla 72. Porcentajes de reseñas nativas de restaurantes vs. Hoteles. ....	212

## Índice de ilustraciones

Ilustración 1. Actividad en Internet (% de personas entre 16 – 74 años, 2022) (Eurostat, 2023: 30). .....	24
Ilustración 2. Usuarios de Internet en Europa que hicieron alguna compra en internet en los doce meses anteriores a las encuestas en 2017 y en 2022 (Eurostat, 2022). .....	25
Ilustración 3. Equipamiento TIC en los hogares españoles de 2012 a 2022 (INE, 2022). .....	27
Ilustración 4. Rasgos discursivos asociados a estrategias de acercamiento al lector (Vásquez, 2015: 35). .....	43
Ilustración 5. Evolución cronológica de la TA (Elaboración propia). .....	59
Ilustración 6. Funcionamiento de Apertium. ....	70
Ilustración 7. Codificación de la oración “My flight is delayed”. Extraída de Forcada (2017). ...	87
Ilustración 8. Decodificación de inglés a español de la oración “My flight is delayed”. Extraída de Forcada (2017). .....	87
Ilustración 9. Estructuras de una RNN y de una BRNN. ....	89
Ilustración 10. Modelo de TAN basado en codificador-decodificador. Extraído de Peris-Abril, Domingo-Ballester y Casacuberta (2017: 6), siguiendo el modelo de Bahdanau, Cho y Bengio, 2014.....	90
Ilustración 11. Representación gráfica de una red neuronal convolucional. Extraído de Cho, Marriënboer, Bahdanau y Bengio, 2014. ....	91
Ilustración 12. Modelo de TAN de Google. Extraído de Wu et al. (2016: 4). ....	93
Ilustración 13. Estructura del modelo Transformer según Vaswani et al. (2017: 3).....	95
Ilustración 14. «Cerebro mecánico» de Artsrouni. Imagen extraída de Hutchins, 2004: 13. ...	105
Ilustración 15. «Máquina de traducir» de Trojanskij. Imagen extraída de Hutchins, 2004: 18.105	
Ilustración 16. Ejemplo de entrada de reseña de hotel. ....	127
Ilustración 17. Ejemplo de entrada de reseña de hotel. ....	127
Ilustración 18. Ejemplo de entrada de reseña de restaurante. ....	128
Ilustración 19. Ejemplo de entrada de reseña de restaurante. ....	128
Ilustración 20. Datos sobre porcentaje de reseñas sobre restaurantes y los principales tipos de errores de localización. ....	140
Ilustración 21. Niveles de calidad de las reseñas sobre restaurantes traducidas con TA. ....	143
Ilustración 22. Porcentaje de errores lingüísticos que afectan a la calidad de la TA. Corpus restaurantes. ....	145
Ilustración 23. Porcentaje de errores culturales que afectan a la calidad de la TA. Corpus restaurantes. ....	148
Ilustración 24. Porcentaje factores que afectan al nivel de calidad de la TA. Corpus restaurantes. ....	150
Ilustración 25. Datos sobre el porcentaje de reseñas sobre hoteles y los principales tipos de errores de localización. ....	155
Ilustración 26. Niveles de calidad de las reseñas sobre hoteles traducidas con TA. ....	159
Ilustración 27. Porcentaje de errores lingüísticos que afectan a la calidad de la TA. Corpus hoteles.....	161
Ilustración 28. Porcentaje de errores culturales que afectan a la calidad de la TA. Corpus hoteles.....	167
Ilustración 29. Porcentaje factores que afectan al nivel de calidad de la TA. Corpus hoteles. ....	173
Ilustración 30. Porcentajes de reseñas según su longitud. Corpus hoteles.....	209
Ilustración 31. Porcentajes de reseñas según su calidad y longitud. Corpus hoteles.....	209
Ilustración 32. Porcentajes de reseñas según su longitud. Corpus restaurantes. ....	210
Ilustración 33. Porcentajes de reseñas según su calidad y longitud. Corpus restaurantes. ....	211

## Índice de ecuaciones

1. Ecuación 1. Representación de la probabilidad de que una oración en LM sea la traducción de una oración en LO. Extraído de Wu et al. (2016: 3)..... 86
2. Ecuación 2. Representación de la probabilidad de que una oración en LM sea la traducción de una oración en LO. Extraído de Kalchbrenner et al. (2016)..... 86

## 1. Introducción

La presente tesis se inspira en que, gracias a la aparición y al desarrollo de Internet, cada vez existen más contenidos en la web, que deben llegar a consumidores de distintas partes del mundo con culturas y lenguas muy distintas. De esta manera, la traducción, la localización y la posesión de dichos contenidos se tornan más necesarias que nunca. El contenido generado por el usuario (CGU) no es una excepción y, al igual que los sitios web o los blogs, estos contenidos necesitan ser localizados. Como en otros sectores, en el ámbito del turismo y la restauración, el CGU ayuda en la toma de decisiones de los consumidores antes, durante y después de su experiencia, por lo que una incorrecta gestión de estos contenidos puede llegar a ser problemática.

El CGU en los ámbitos del turismo y la restauración se publica normalmente en portales o plataformas de reseñas como TripAdvisor y, debido a las grandes cantidades de CGU publicado en estas plataformas, habitualmente, se opta por utilizar mecanismos de traducción automática, lo que puede dar lugar a errores si no hay un traductor humano que posea estas traducciones. Es más, si el CGU no se localiza correctamente, puede ocasionar importantes pérdidas de contenido, ya que el cliente final podría no llegar a entender lo que intenta comprar y, por tanto, decidir no adquirir dicho producto o servicio. De esta manera, el CGU y su correcta gestión y adaptación se convierten en partes esenciales en la estrategia de marketing.

Este estudio está motivado por mi inclinación hacia los aspectos más técnicos de la traducción, es decir, mi interés en la traducción automática, las herramientas de traducción asistida por ordenador, la localización o la posesión de textos traducidos con traducción automática. Además, desde la realización de mi máster en Localización y la elaboración de mi trabajo de final de máster sobre este mismo tema, entendí la gran importancia que esta disciplina tiene en los procesos de internacionalización de las empresas. Asimismo, gracias a mis experiencias de trabajo en campos como el marketing y el comercio internacional, descubrí el vital papel que juegan los contenidos correctamente localizados en las estrategias empresariales.

Con este propósito, se ha realizado un recorrido por la historia del CGU y se han recogido las diferentes definiciones de los autores que han hablado de CGU desde la aparición de este hasta la actualidad. De la misma manera, se ha retomado el estudio que realicé en mi proyecto de fin de máster (2018) donde se abordaba la localización, su historia, la actualidad y las perspectivas de futuro, para recalcar la importancia de la localización del CGU. Además, también se estudia a fondo la TA, observando su historia y los tipos de TA desde su aparición hasta la actualidad.

La necesidad de este estudio radica en la escasa bibliografía referente a este tema. Si se habla de traducción y turismo, es posible encontrar miles de artículos, que abordan la temática del lenguaje del turismo, de las técnicas de traducción, de la terminología turística, etc. De igual manera, también existen muchos estudios donde se investigan los diferentes paradigmas de TA, para comprender cómo funciona, cómo ha ido evolucionando e incluso los nuevos trabajos y retos que presenta en la actualidad para los traductores profesionales. Sin embargo, la literatura sobre la TA de contenido generado por el usuario en el campo del turismo es más escasa.

Por último, se han examinado también diversa bibliografía donde se aborda el CGU en diferentes ámbitos (turismo, comercio electrónico, redes sociales, periodismo, ingeniería, informática, etc.). En estos trabajos se proporcionan definiciones de lo que este representa y se recalca su importancia. No obstante, no se ha localizado ningún trabajo donde se conecten las variables mencionadas en este trabajo a través de la recolección y el análisis de un corpus formado por CGU en plataformas de reseñas en línea de turismo y restauración como, por ejemplo, TripAdvisor. El presente estudio pretende, por lo tanto, dar una definición propia del CGU, pero también aportar una visión objetiva sobre la importancia de una correcta estrategia de posesición y localización. Además, también nos permitirá arrojar luz sobre las potenciales posibilidades de futuro de los profesionales de la traducción.

### **1.1. Premisas de trabajo**

La presente investigación parte de las siguientes premisas:

1. El contenido generado por el usuario (CGU) publicado en plataformas de reseñas como TripAdvisor puede afectar a la percepción del producto o servicio turístico por parte de los consumidores de dicho contenido.
2. Si el CGU no se traduce, la información se pierde, y esto, a su vez, afecta de manera directa a las ventas de dichos productos o servicios. Por este motivo, tanto el CGU como su correcta gestión se convierten en parte clave de las estrategias de marketing e internacionalización de las empresas.
3. Si el CGU se traduce mediante traducción automática (TA) se pueden obviar aspectos intrínsecos de cada cultura. Es decir, si el contenido no llega localizado al mercado meta, los consumidores de dicho mercado no podrán entenderlo o interpretarlo. Como resultado, puede que opten por no consumir determinado producto o servicio. Por este motivo, se considera que la localización de los contenidos publicados en internet es necesaria.

4. La TA ayuda en este tipo de contextos, pero sigue siendo necesaria la posesición por parte de un traductor humano y nativo de la lengua meta que pueda revisar los textos traducidos con motores de TA.
5. La longitud de las reseñas afecta a la calidad de los resultados de TA, cuanto más larga es la reseña, peor será la calidad de los contenidos traducidos.

## 1.2. Objetivos

El objetivo principal de este trabajo es definir el CGU y aportar una visión objetiva sobre la importancia de implementar una correcta estrategia de posesición que tenga en cuenta los principios de la localización y las métricas de evaluación de la calidad de los contenidos traducidos con TA. Toda buena estrategia de marketing debería incluir una producción de textos enfocada a la cultura destinataria, por consiguiente, es imperativo que las traducciones sean correctas y que el usuario entienda el contenido que se le presenta. Si el consumidor no sabe o no entiende qué está comprando, probablemente no se decida por dicho producto o servicio.

De esta manera, se han planteado los siguientes objetivos específicos:

1. Analizar el género textual de las reseñas en plataformas en línea.
2. Comprobar la validez de las métricas tradicionales para los nuevos sistemas de TA neuronal.
3. Estudiar cómo afecta la longitud de las reseñas a la calidad de los resultados de TA neuronal.

La investigación servirá para detectar posibles necesidades de localización y posesición en este género textual, cuando se traduce mediante TA. Además, también será posible extraer conclusiones sobre cómo la localización de dichos textos se convierte en una herramienta crucial de la estrategia de marketing e internacionalización empresarial.

De esta manera, la investigación sentará unas bases para futuros estudios que permitirán identificar nuevas formas y métodos de posesición de textos traducidos con TA siguiendo los principios de localización, así como descubrir qué nuevas posibilidades ofrece dicha variedad de traducción a los futuros traductores.

## 1.3. Estructura del trabajo

La primera parte del estudio se dedica al marco teórico de la traducción automática, el contenido generado por el usuario y la localización, además, de los contextos en que estos se emplean y su papel en la Web 2.0 y el *Big Data*. Para la elaboración de esta sección, se recurre a publicaciones de varios tipos (desde artículos en revistas especializadas hasta tesis doctorales) publicadas desde los inicios de la traducción automática hasta el presente. Estos artículos provienen tanto del mundo académico (tesis doctorales, libros, capítulos en libros, etc.) como del mundo profesional (artículos o estudios en revistas especializadas del sector). Para el análisis

de dicha bibliografía y los antecedentes, se han utilizado estudios y publicaciones extraídos de BITRA (base de datos de estudios de traducción e interpretación de la Universidad de Alicante) y TESEO (base de datos de tesis doctorales del Ministerio de Educación).

La parte central de esta investigación consiste en el estudio de un corpus de textos de reseñas de hoteles y restaurantes, escritas originalmente por hablantes nativos de inglés, en la plataforma de reseñas especializadas de turismo, TripAdvisor, plataforma escogida por el carácter local de las reseñas, la gran variedad de CGU de diferentes tipos (foros, *chartooms*, reseñas, valoraciones, ideas y experiencias de viajeros) que se publican en ella y por ser una de las más utilizadas hoy en día con aproximadamente 73 millones de reseñas y opiniones en 2022 (TripAdvisor Transparency Report, 2023). Para realizar un estudio contrastivo y observar en qué difieren las reseñas traducidas mediante traducción automática (escritas originariamente en inglés), de las reseñas escritas por hablantes nativos de español, se ha recopilado un corpus de 500 reseñas, 250 reseñas de hoteles y 250 reseñas de restaurantes de comida típica valenciana, escritas originalmente en inglés.

El motivo por el cual se ha dividido el corpus y el análisis en reseñas de hoteles y reseñas de restaurantes es que cada una de estas reseñas cuenta con características diferentes, que determinarán la forma en que los motores de TA las procesan. Por ejemplo, como se comprobará más adelante, las reseñas de restaurantes cuentan con mucho más vocabulario con altas cargas culturales y las reseñas de hoteles suelen ser más largas, ambos factores determinantes de la calidad de las reseñas traducidas con TA. Para reunir y estudiar dichos corpus de reseñas, se han utilizado, entre otras herramientas informáticas, WordSmith para extracción de patrones y ocurrencias en el corpus, así como Microsoft Access y Microsoft Excel para la elaboración y el análisis de la base de datos.

Las 500 reseñas se clasificaron en cuatro niveles de calidad basados en el modelo de análisis de la calidad DQF de TAUS (Van der Meer et al., 2017): “nativas”, cuando el texto no contiene errores de ningún tipo y se asemeja a un texto que escribiría un hablante nativo de Español (de España en este caso); “correctas”, cuando la reseña no contiene errores notables, pero la expresión suena forzada y se puede percibir que no fue escrita por un hablante nativo; “aceptables”, si se encuentran errores, pero la reseña todavía transmite el significado, por lo que el usuario puede llegar a utilizarla para entender el contenido; y, por último, “inaceptables”, que son aquellas reseñas con numerosos errores que no permiten a los usuarios entender el mensaje. Pese a que esta clasificación está basada en el modelo de análisis de la calidad DQF de TAUS, en este estudio, la clasificación se realiza teniendo en cuenta la calidad de las reseñas



completas y no teniendo en cuenta el nivel de gravedad de los errores individuales como propone TAUS (Van der Meer et al., 2017).

Para ilustrar mejor los niveles de calidad de cada categoría se ha elaborado la Tabla 1, a continuación:

Tabla 1. Ejemplos de clasificación de las reseñas (Elaboración propia).

Original EN	TA ES	Nivel de calidad
<i>The interior of the restaurant is very pleasant.</i>	El interior del restaurante es muy agradable.	Nativa
<i>A very nice place where to eat paella and drink sangria - and also very cheap: we paid 30€ for a fixed menu and every dish was made with very fresh ingredients.</i>	Un lugar muy agradable donde comer paella y beber sangría, y también muy barato: pagamos 30 € por un menú fijo y cada plato estaba hecho con ingredientes muy frescos.	Correcta
<i>We had a very complete breakfast too.</i>	Tuvimos un desayuno muy completo también.	Aceptable
<i>Great breakfast, advise you would like GF bread/cakes and the lovely friendly lady in charge of dining room looks after you for your stay.</i>	Gran desayuno, te aconsejo que quieras pan / pasteles GF y la encantadora y amable señora a cargo del comedor te cuida durante tu estancia.	Inaceptable

Una vez compilado el corpus, se han llevado a cabo diferentes análisis. Primero, se realizó un análisis sobre las problemáticas de localización que presentan las reseñas de hoteles y de restaurantes traducidas mediante TA, para lo que se utilizaron los dos corpus bilingües (el de reseñas de hoteles y el de restaurantes). Estos corpus cuentan con 500 reseñas escritas en inglés extraídas de Tripadvisor.com, que se han procesado con el motor de TA de la plataforma, el Traductor Google, para obtener su traducción automática al español. Las reseñas resultantes, se han analizado teniendo en cuenta los principios de la localización, con lo que se han identificado los errores que afectan a la calidad de las reseñas traducidas con TA, a nivel lingüístico y cultural.

Así, el primer estudio llevado a cabo ha sido un análisis cuantitativo y cualitativo de los corpus bilingües de reseñas traducidas al español con TA. En dicho estudio, se ha examinado el número de errores de localización en cada uno de los corpus y se ha analizado cómo estos afectan a la calidad de las traducciones. El segundo estudio ha sido un análisis cualitativo en que se han

estudiado los problemas lingüísticos y culturales de localización, así como las causas frecuentes que influyen en el nivel de calidad de las reseñas traducidas con TA.

El tercer estudio, ha sido un análisis de frecuencias que ha servido para determinar hasta qué punto difieren las reseñas traducidas al español con TAN de las reseñas redactadas por hablantes nativos de español (en este caso se ha centrado la investigación en el español de España por ser la autora procedente de este país) en esta lengua. En este caso se han compilado dos corpus monolingües, consistentes en 500 reseñas (250 de restaurantes y 250 de hoteles) extraídas de Tripadvisor.com escritas en español por hablantes nativos de español de España. Tras procesar los corpus, se han obtenido listas de frecuencias que nos han permitido realizar una comparativa sobre la naturalidad y la autenticidad de los corpus monolingües y las de los corpus bilingües.

El cuarto y último análisis, se ha centrado en determinar cómo afecta la longitud de las reseñas al nivel de calidad de las reseñas traducidas con TAN. Para ello primero se ha calculado la longitud media de las reseñas de cada corpus. Después se han calculado los porcentajes correspondientes a las reseñas largas y a las reseñas cortas y, finalmente, se han calculado los porcentajes de reseñas nativas, correctas, aceptables e inaceptables según si estas son largas o cortas.

Finalmente, se ha observado cómo la corrección o incorrección del CGU, así como su adecuación y adaptación (localización) al mercado o *locale* meta, podrían influir en el comportamiento de los consumidores y en su interacción con el producto o servicio. Con la lectura y el análisis de estos estudios, se obtendrán datos significativos para extraer conclusiones sobre el desarrollo futuro de la traducción y para plantear estrategias de localización y posesición adecuadas.

Con todo esto, la tesis se divide en cinco capítulos dedicados a presentar el estado actual de la traducción automática de contenidos generados por el usuario, en concreto, de reseñas de hoteles y restaurantes en portales de turismo en línea, para dar cuenta de la calidad de dichos contenidos traducidos con motores de traducción automática.

El primer capítulo, la introducción, cuenta con seis apartados. En primer lugar, se presentan las premisas y los objetivos sobre los cuales se ha trabajado. Después, se profundiza en la motivación para llevar a cabo esta tesis y se hace hincapié en la necesidad de estudio, la relevancia del tema y el estado en que se encuentra la investigación en esta área de estudio. A continuación, se describe la metodología seguida para llevar a cabo los análisis empíricos sobre la calidad de la traducción automática de las reseñas.

El segundo capítulo introduce el marco teórico para las diferentes temáticas tratadas y cuenta con tres apartados con sus correspondientes subapartados. El primer apartado trata sobre contenido generado por el usuario. Este, a su vez, cuenta con cinco subapartados donde se detalla la aparición y el desarrollo del CGU, desde sus inicios fuera de internet (*Word-of-mouth*), hasta la Web 2.0, donde los usuarios ya no solo consumen el contenido, sino que también lo crean. También se ofrecen las diferentes definiciones que se han proporcionado hasta el momento, así como las principales características del CGU y la importancia de estudio de esta temática. Además, también se exponen los diferentes puntos de vista desde los que se ha estudiado este tipo de contenido, así como las ventajas que este presenta para las empresas si se gestiona correctamente. Finalmente, se presentan los diferentes tipos de CGU que existen en la actualidad, entre los cuales se encuentran las reseñas generadas por el usuario, que son el objeto de estudio de esta tesis.

El segundo apartado trata sobre la localización y cuenta con cuatro subapartados, en que se precisan sus orígenes y se presentan las diferentes definiciones de este concepto, desde su aparición sobre los años 80 hasta la actualidad. También se detallan las características y los fundamentos de esta disciplina y, además, se exponen aquellos rasgos que la diferencian de la traducción.

El tercer apartado, dedicado a la traducción automática, es el más extenso del marco teórico y cuenta con seis subapartados que presentan una visión muy detallada de la aparición, el desarrollo y el estado actual de la traducción automática. Primero, se han presentado algunas de las principales definiciones de este concepto. A continuación, se han detallado los diversos tipos de mecanismos de traducción automática que han surgido a lo largo de los tiempos, desde la traducción automática basada en reglas, pasando por la traducción automática basada en corpus (estadística y basada en ejemplos), la traducción automática híbrida y hasta llegar a uno de los modelos más recientes y utilizados en los últimos años, el de traducción automática neuronal (TAN). Dada la relevancia que ha cobrado este modelo en los últimos años, gracias a motores como el de Google (*Google Neural Machine Translation* o GNMT), se han dedicado varios subapartados sobre la TAN. En dichos subapartados, se exponen la definición de TAN, las características y el funcionamiento, así como, una introducción al motor neuronal de Google. Además, se habla del modelo *transformer*, uno de los más recientes e innovadores, y que está cobrando gran relevancia actualmente dados los resultados que ofrece. Seguidamente, se detallan las ventajas e inconvenientes de utilizar la TAN y se hace una reflexión final sobre la TAN.

Después de presentar los diferentes tipos de traducción automática, se dedica un subapartado a la posesición, donde se define esta disciplina, se comentan sus características y los puntos de vista desde los que se ha estudiado. Además, se alude a su gran impacto sobre el papel del traductor profesional actual y se menciona el escepticismo que algunos profesionales y académicos todavía sienten hacia esta disciplina. Posteriormente, en el subapartado de las métricas de evaluación de la calidad, se presentan los principales modelos utilizados para medir la calidad de la TA (modelo DQF de TAUS). Además, se menciona la falta de unificación de estas métricas y de los parámetros utilizados por estas para medir la calidad de la TAN.

Tras este detallado subapartado sobre el estado de la TA, se ha hecho referencia a los orígenes de la traducción automática. En este subapartado, se detalla la historia de la TA desde su inicio sobre los años 30, con Artsrouni y Trojanskij, y se menciona la influencia de Alan Turing y la criptografía, los cuales tuvieron gran influencia en el desarrollo de los ordenadores y en las primeras teorías sobre inteligencia artificial. También se ha incluido información sobre el informe ALPAC y sus consecuencias durante la guerra fría (años 60-70), así como sobre la posterior comercialización sobre los sistemas de TA y las primeras herramientas traducción asistida por ordenador (herramientas TAO). Por último, se mencionan los desarrollos e innovaciones aparecidos desde los años 2000 hasta la actualidad y se hace una reflexión de cómo ha evolucionado la TA según la literatura estudiada.

El tercer capítulo presenta los diferentes análisis y resultados llevados a cabo en esta tesis. Este capítulo cuenta con ocho apartados correspondientes a los diferentes análisis llevados a cabo y a las observaciones sobre estos. El capítulo empieza con la descripción sobre el diseño del estudio y un breve recordatorio de la metodología utilizada para los diferentes análisis. El primer análisis corresponde al análisis cualitativo y cuantitativo de las reseñas de hoteles y de las reseñas de restaurantes y en este, se detalla el número de errores encontrados y los porcentajes de reseñas nativas, correctas, aceptables e incorrectas totales de cada corpus de reseñas traducidas con TAN.

En el segundo análisis, se lleva a cabo un análisis que compara la naturalidad de las reseñas. Para ello, se han extraído las palabras más frecuentes de las reseñas escritas por hablantes nativos de español (corpus monolingüe) y las palabras más frecuentes de las reseñas traducidas con TAN (corpus bilingüe). Así, se han observado las características diferentes que presentan las reseñas de estos dos corpus. Por último, en el tercer análisis, se realiza un estudio que determina si la longitud de las reseñas influye en la calidad de los resultados de TAN. En este apartado, se detallan los parámetros utilizados para el análisis de la longitud, se obtiene la media de los

porcentajes de reseñas largas y cortas (extraído según la bibliografía consultada sobre este tema) y se calculan los porcentajes de reseñas según su longitud y su nivel de calidad.

El cuarto capítulo consiste en las conclusiones de esta tesis. En él, se recapitula el contenido y se presentan las principales conclusiones basadas en el análisis y los resultados obtenidos. Además, también se presenta un listado de posibles futuras líneas de estudio en este ámbito. En último lugar, en el quinto capítulo, se ha añadido la bibliografía consultada y, tras esta, los anexos que presentan todos los materiales con los que se ha trabajado.

## 2. Marco Teórico

En este apartado, se presenta el marco teórico en que se ha basado esta tesis. Primero, se ha explorado el contenido generado por el usuario (CGU) desde sus orígenes, hasta su evolución a lo que es hoy en día. Además, se han presentado sus principales características y ventajas, así como los diferentes tipos de CGU existentes en la actualidad. Posteriormente, se ha abordado el concepto de Localización, así como su historia, características y sus principales definiciones, entre otros temas. Por último, se ha realizado una exhaustiva revisión bibliográfica de la traducción automática, desde sus orígenes, hasta los modelos más novedosos en la actualidad. Asimismo, se abordan los conceptos de posesión y de métricas de calidad de traducción automática.

### 2.1. Contenido Generado por el Usuario (CGU)

En el presente apartado de la tesis, se realizará un recorrido por los orígenes del contenido generado por el usuario (CGU), desde el boca-a boca tradicional o *word-of-mouth*, la aparición de la Web 2.0, hasta la evolución del CGU hacia diferentes tipos de contenido como las reseñas generadas por el usuario (RGU). Para la presente tesis, se utilizarán las denominaciones «contenido generado por el usuario» (CGU) y «reseñas generadas por el usuario» (RGU) para referirse a cada uno de estos conceptos que se desarrollarán a lo largo del trabajo. Además, se analizará su posible influencia sobre los comportamientos de compra de los usuarios y se observarán las numerosas ventajas que ofrece este contenido tanto a los usuarios como a las empresas proveedoras de productos y servicios turísticos y de restauración, cuando este se posea adecuadamente tras ser traducido mediante sistemas de traducción automática.

#### 2.1.1. Antecedentes

Para entender cómo el CGU ha evolucionado hasta llegar a lo que es hoy en día, primero es necesario mencionar el boca-a-boca tradicional o *word-of-mouth* (WOM por sus siglas en inglés). El CGU como tal no existía en formato escrito puesto que se solía transmitir de manera oral (WOM) y sin una estructura específica. Además, por su carácter oral, el WOM tradicional llega a una audiencia mucho más reducida y local (Dellarocas et al., 2004: 03). Las tecnologías digitales han dado lugar a nuevas formas de comunicarse y de relacionarse, lo que está suponiendo un reto para nuestra concepción de las amistades, el trabajo, el ocio, la salud, el aprendizaje o la política, entre otros ámbitos (Jones, Chik y Hafner, 2015: 01).

Entre estas nuevas formas de comunicación, aparecen algunos espacios interactivos de escritura como los blogs y otras redes sociales, donde se dan múltiples formas de interacción social muy

distintas de aquellas que se pueden encontrar en las conversaciones cara a cara y en los textos escritos tradicionales (Jones, Chik y Hafner, 2015: 01). Así, con la aparición de estas nuevas tecnologías y redes sociales, los usuarios pueden digitalizar sus experiencias (por ejemplo, cuando suben videos o fotos), lo que ha creado un nuevo paradigma de la manera en que se comparte la información, el cual se ha denominado «boca-a-boca electrónico» o *electronic word-of-mouth* (eWOM) (Manap y Adzharudin, 2013; Ruzic y Biloz, 2010; Sparks y Browning, 2010).

De esta manera, el WOM, tradicionalmente oral, evoluciona hacia el eWOM cuando este se extiende a las redes sociales, blogs y demás plataformas de Internet. El eWOM, al contrario que el WOM, se caracteriza por la escalabilidad y la rapidez de difusión, además llegar a muchos más usuarios (Dellarocas et al., 2004: 03). Una de las principales diferencias entre ambos es que el WOM tradicional es efímero, puesto que es oral y acaba desvaneciéndose, pero el eWOM, al transmitirse de manera escrita, siempre queda registrado. Además, el eWOM se produce en un entorno digital en el que el autor del CGU y el usuario que lo consume posiblemente no llegarán a conocerse entre sí, algo que es imposible en el WOM tradicional. En palabras de Vásquez (2015: 21):

It takes place in a technologically mediated form, between an author and a potentially vast audience, both of whom are – and will most likely remain – unknown to one another in an offline sense. At the present time, online consumer reviews are a predominantly text-based asynchronous (and often, anonymous) genre of computer mediated communication, or CMC.

De esto se puede extraer que las principales características que se encuentran en el nuevo paradigma del eWOM son la gran escalabilidad y alcance del contenido, la mayor duración en el tiempo de estos contenidos por su carácter escrito y, por supuesto, que este se produce en el ámbito digital, el cual no es homogéneo y puede adoptar tipologías, formas y géneros distintos.

Hennig-Thurau et al. (2004: 39) definen el eWOM como «any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet». Es decir, que identifican el eWOM como cualquier contenido o valoración (negativa o positiva) de Internet que producen los consumidores para que otros consumidores lo lean y lo puedan usar para informarse. Vásquez (2015: 20) añade que el eWOM es la forma de discurso corporativo o comercial generado por el usuario más generalizado y dominante que existe hoy en día.

Dada la cantidad emergente de comentarios y reseñas publicados en Internet por los usuarios en los últimos años, el CGU podría considerarse como un nuevo género digital (Candel-Mora, 2015; Pollach, 2006; Vásquez, 2015). No obstante, en sus inicios, el CGU aparece en otros medios no digitales, así, se puede encontrar el contenido generado por el lector (CGL) o el radioyente (CGR) o, en general, el contenido generado por la audiencia (CGA) (García de Torres, 2010: 586). Según García de Torres (2010: 586), el CGU en sus primeros años de presencia online está presente en grupos de noticias, foros, tableros de anuncios, páginas web editadas por usuarios y chats, formatos con bastantes limitaciones.

Más tarde, aparecieron plataformas que ampliaron las posibilidades de creación y distribución del CGU, lo que hizo patente el potencial de transformación de la estructura en la que se abrió paso la Web 2.0 (García de Torres, 2010: 586; Hidalgo-Alcázar, Sicilia y Ruiz de Maya, 2015: 80). Candel-Mora (2015: 40) añade que han aparecido nuevos géneros digitales gracias a la evolución de las tecnologías de la información y de la comunicación, y ejemplifica esta evolución con el surgimiento de las redes sociales y la participación de los usuarios en la Web 2.0.

En la Web 2.0, los consumidores cobran un mayor protagonismo y pasan de ser meros espectadores de lo que ven y leen en Internet a convertirse en creadores de contenidos y servicios, es decir, tienen el control sobre el contenido de la web (Cebrián, 2008: 349; Nafría, 2007). Cebrián (2008: 347) define la Web 2.0 como una plataforma de redes sociales de información donde los datos están siempre disponibles para cualquier persona en cualquier lugar con acceso a Internet. En resumen, la Web 2.0 podría definirse como el conjunto de sitios web y plataformas en línea que permiten que los usuarios puedan crear contenidos para compartir información y colaborar en Internet. Es decir, es un espacio colaborativo e interactivo, donde los usuarios actúan como consumidores, pero también como creadores de contenido en el entorno digital de Internet.

Esta evolución en el perfil de los consumidores propicia que los modelos de información también cambien. Así, los modelos verticales que había hasta el momento se transforman en modelos horizontales, en los que no hay orden jerárquico entre los usuarios, sino que todos ellos están al mismo nivel, sin prioridades (Cebrián, 2008: 349). Los consumidores han pasado, por tanto, a ser prosumidores, es decir, ya no solo se dedican a adquirir los productos o servicios publicitados en las redes, sino que además producen contenidos en Internet que, a menudo terminan siendo leídos por otros consumidores (Muller et al., 2011: 24).



A continuación, la Ilustración 1 muestra el principal uso de Internet que realizan los ciudadanos de la Unión Europea (UE) según un informe del Eurostat de Strandell y Wolff (2019):

## Internet activities

(%, share of people aged 16–74 years, 2022)

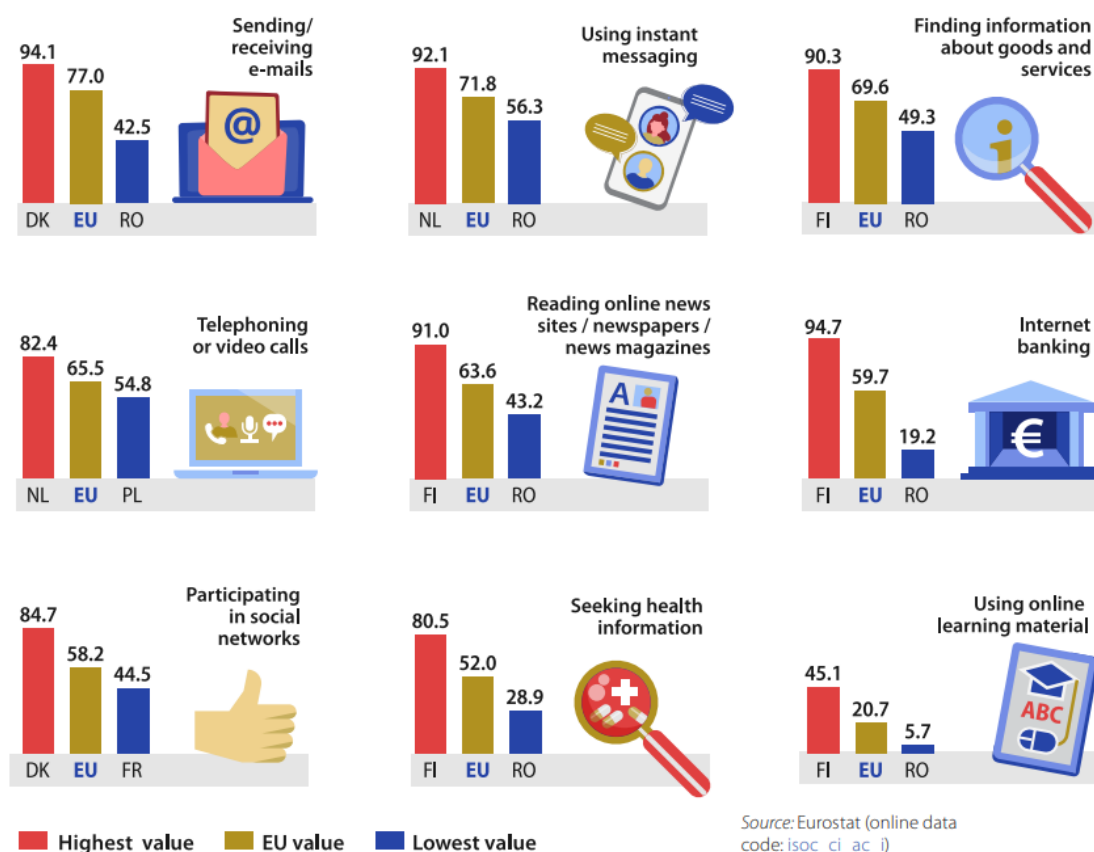


Ilustración 1. Actividad en Internet (% de personas entre 16 – 74 años, 2022) (Eurostat, 2023: 30).<sup>1</sup>

Como se observa en la Ilustración 1, la columna azul muestra los usuarios de Internet procedentes de los 28 países europeos (ahora 27 con la salida de Gran Bretaña) de entre 16 y 74 años<sup>2</sup> que utilizan Internet con diversos propósitos. Entre dichos propósitos se encuentran enviar y recibir correos electrónicos (77 %), utilizar mensajería instantánea (71.8 %), encontrar información sobre bienes y servicios (69.6 %), realizar llamadas y videollamadas (65.5 %), leer contenidos digitales (63.6 %), gestionar sus finanzas (59.7 %), participar en las redes sociales (58.2 %), buscar información médica (52 %), y cursos y educación en Internet (20.7 %). En gran

<sup>1</sup> Eurostat (2023). Key figures on Europe 2023 edition. [Data set]. <https://ec.europa.eu/eurostat/documents/15216629/17706595/KS-EI-23-001-EN-N.pdf/5df7a393-8461-9270-7eaa-91a4b1c2acc6?version=3.0&t=1697616189434>

<sup>2</sup> En el estudio del Eurostat (2023: 30) se utiliza esta franja de edad porque representa el fragmento de población europea más activo en Internet.

parte, la información sobre alojamientos y viajes o sobre bienes y servicios suele ser CGU redactado por otros usuarios.

Además, resulta también interesante observar cómo han evolucionado las compras realizadas por los usuarios a través de internet, ya que, como muestra la Ilustración 2, a continuación, la tendencia desde 2017 se muestra siempre al alza en todos los países de la Unión Europea representados en el gráfico:

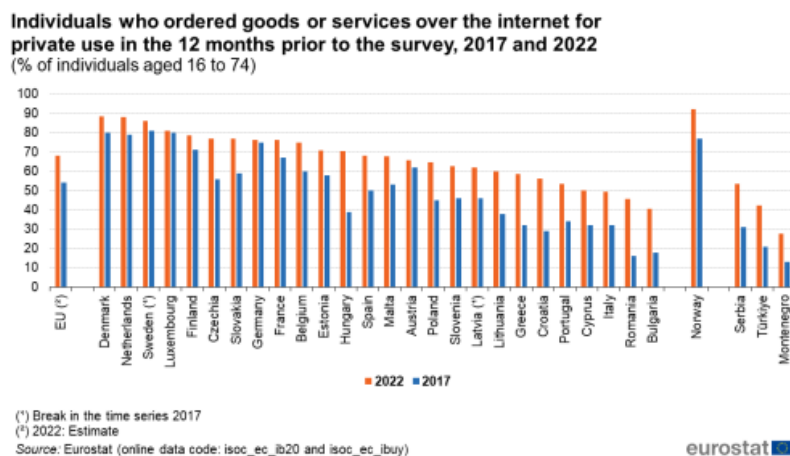


Ilustración 2. Usuarios de Internet en Europa que hicieron alguna compra en internet en los doce meses anteriores a las encuestas en 2017 y en 2022 (Eurostat, 2022).<sup>3</sup>

Como se puede observar en la Ilustración 2, en todos los países se ha dado una acusada tendencia al alza en cuanto a los porcentajes de pedidos realizados a través de internet. En el caso de España, se pasa de un 50 % en 2017 a un 68 % en 2022. Con estos datos, sería posible concluir que, dado el porcentaje cada vez mayor de productos vendidos en internet en toda Europa, las descripciones de estos deberían adaptarse de manera adecuada a los usuarios de estos países, si no estas ventas podrían verse afectadas.

Estas cifras se acusan todavía más si se centra la atención solamente en España, como muestra la Ilustración 3 siguiente extraída de un informe del Instituto Nacional de Estadística (INE) sobre las diferentes actividades que realizaron los usuarios españoles en Internet en 2022 (INE, 2022):

<sup>3</sup> Eurostat (2022). Individuals who ordered goods or services over the internet for private use in the 12 months prior to the survey, 2017 and 2022. [Data set]. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Individuals\\_who\\_ordered\\_goods\\_or\\_services\\_over\\_the\\_internet\\_for\\_private\\_use\\_in\\_the\\_12\\_months\\_prior\\_to\\_the\\_survey,\\_2017\\_and\\_2022\\_\(%25\\_of\\_individuals\\_aged\\_16\\_to\\_74\)\\_14-12-2022.png](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Individuals_who_ordered_goods_or_services_over_the_internet_for_private_use_in_the_12_months_prior_to_the_survey,_2017_and_2022_(%25_of_individuals_aged_16_to_74)_14-12-2022.png)

Tabla 2. Porcentaje de usuarios de Internet en los últimos 3 meses por tipo de actividad realizada. 2022 (% de personas de 16 a 74 años).<sup>4</sup>

Porcentaje de usuarios de Internet en los últimos 3 meses por tipo de actividad realizada. 2022 (% de personas de 16 a 74 años)		
	Mujeres	Hombres
Usar mensajería instantánea (vía WhatsApp, Skype, Messenger...)	89,7	82,9
Telefonar o recibir videollamadas a través de internet	61,9	58,7
Leer noticias, periódicos, revistas online	53,0	68,3
Recibir o enviar correos electrónicos	45,4	66,6
Buscar información sobre bienes y servicios	44,2	54,4
Buscar información sobre temas de salud	39,5	42,3
Utilizar banca por internet (incluida banca móvil)	35,4	54,3
Participar en redes sociales	32,2	26,4
Concertar una cita médica por web o app	30,1	35,8
Emitir opiniones sobre asuntos de tipo cívico o político	11,8	6,5
Tomar parte en consultas online o votaciones sobre asuntos cívicos o políticos	11,0	9,3
Utilizar material de aprendizaje on line que no sea un curso completo on line	6,2	6,5
Vender bienes o servicios a través de un sitio web o app	5,6	5,6
Realizar algún curso on line	3,3	7,3

Según los datos reflejados en la Tabla 2, el principal uso de las tecnologías de internet por parte de los españoles de entre 16 y 74 años (en porcentaje de mujeres y hombres) es el de la mensajería instantánea (89,7 % y 82,9 % respectivamente), seguido de telefonar o realizar videollamadas a través de internet (61,9 % y 58,7 % respectivamente). El porcentaje de personas que buscan información sobre bienes o servicios se encuentra en el 44,2 % para las mujeres, y el 54,4 % para los hombres. También se observa que, siguiendo la tendencia general de los países europeos, los usos relacionados con la recepción y el envío de emails (45,4 % y 66,6 %), la banca electrónica (35,4 % y 54,3 %) o la lectura de noticias (53 % y 68,3 %) también son de los principales en España. En relación con el CGU, se han podido encontrar los datos correspondientes al número de personas que participan en redes sociales (32,2 % y 26,4 %) y que emiten opiniones sobre asuntos sociales o políticos (11,8 % y 6,5 %).

Estos datos demuestran lo presente que está el CGU en el día a día del consumo de Internet de los usuarios y permiten ver cómo este contenido evoluciona y en qué medida podría influir en el comportamiento de los consumidores en la web. No obstante, la Ilustración 3 no incluye los datos de consumo de reseñas u opiniones en plataformas destinadas a este fin, tanto en el

<sup>4</sup> INE (2023). Porcentaje de usuarios de Internet en los últimos 3 meses por tipo de actividad realizada. 2022 (% de personas de 16 a 74 años)  
<https://www.ine.es/ss/Satellite?blobcol=urldata&blobheader=Unknown+format&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3B+filename%3DC4T3.xlsx&blobkey=urldata&blobtable=MungoBlobs&blobwhere=279%2F587%2FC4T3%2C0.xlsx&ssbinary=true>

ámbito del turismo como en otros, un dato que parece relevante vista la importancia que dichas reseñas han cobrado en Internet en los últimos años.

Por otra parte, la evolución de la Web 2.0 ha sido posible gracias a la generalización de Internet y de los dispositivos electrónicos en los hogares de los consumidores, lo que se ve claramente en la Ilustración 3 recuperada del Instituto Nacional de Estadística (INE, 2022), la cual muestra el equipamiento de tecnologías de la información (TIC) en los hogares españoles de 2012 a 2022.

### Evolución del equipamiento TIC en los hogares

Porcentaje de hogares con algún miembro de 16 a 74 años

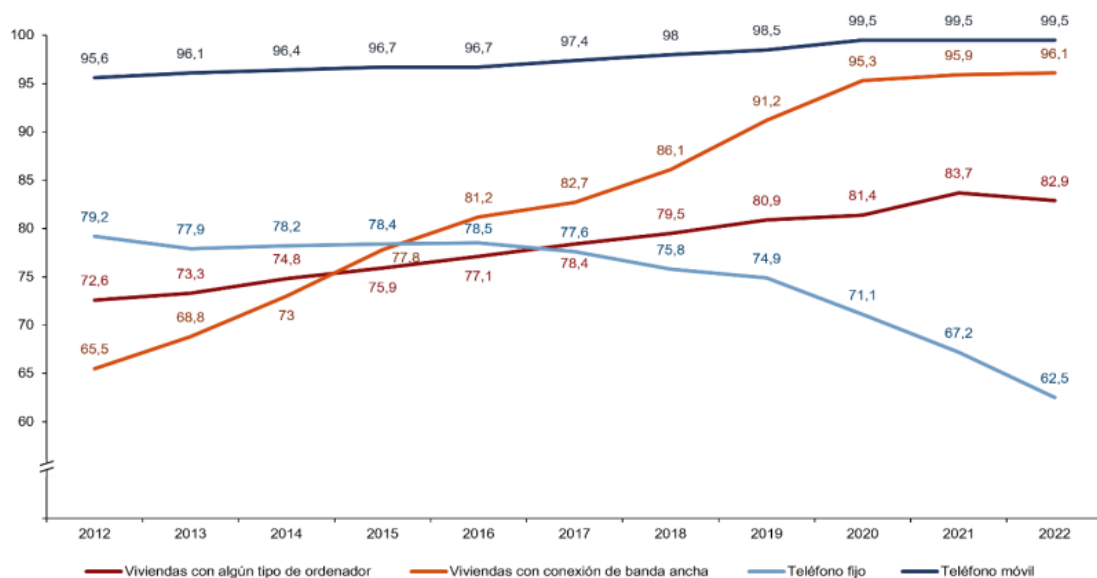


Ilustración 3. Equipamiento TIC en los hogares españoles de 2012 a 2022 (INE, 2022).<sup>5</sup>

En la Ilustración 3, se observa una tendencia al alza desde el año 2012 hasta el año 2022 en el porcentaje de viviendas que disponen de conexión de banda ancha (de un 65,5 % en 2012 a un 96,1 % en 2022) y en el porcentaje de viviendas con teléfono móvil (de un 95,6% en 2012 a 99,5% en 2022). Sin embargo, el porcentaje de viviendas con algún tipo de ordenador ha mostrado un crecimiento más moderado (de un 72,6 % en 2012 a un 83,7 % en 2021) que ha descendido en 2022 en un punto hasta el 82,9 %. Esto a su vez, ha permitido un mayor acceso a los contenidos de Internet en los hogares españoles y por tanto un mayor consumo y producción de CGU.

<sup>5</sup> INE (29 de noviembre de 2022). Equipamiento y uso de TIC en los hogares - Año 2022. [Nota de prensa]. Recuperado de: [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=estadistica\\_C&cid=1254736176741&menu=ultiDatos&idp=1254735976608](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=estadistica_C&cid=1254736176741&menu=ultiDatos&idp=1254735976608)

Pese a esta acusada evolución y expansión de los nuevos medios digitales de comunicación, los géneros digitales se consideran todavía muy nuevos, por lo que todavía no se han asentado sus reglas. Darics (2015: 01) afirma que esto ha ocurrido por dos razones:

1. Las nuevas tecnologías de la comunicación, las cuales cambian y evolucionan constantemente, crean nuevos contextos sociales y comerciales, lo que a su vez genera nuevas situaciones comunicativas. Esto modifica y da forma al lenguaje, al discurso y a las interacciones constantemente.
2. La manera en que la gente utiliza el lenguaje para comunicarse a través de estas plataformas tecnológicas también está en un estado de cambio constante. Todavía no ha dado tiempo a regular y formalizar el discurso en las redes, por lo que, dependiendo del contexto social y tecnológico en línea, dicho discurso cambia (Androutsopoulos, 2006; Darics, 2015: 01; Herring, 2012: 2338).

Como apunta Vásquez (2014: 65), el CGU es una de las características principales de las redes sociales y conlleva que, hoy en día, cualquier persona con acceso a Internet pueda dar su opinión y relatar sus propias experiencias. En conclusión, se podría afirmar que la Web 2.0 ha cambiado la manera en que los usuarios reciben y comparten información. Mientras que anteriormente los consumidores confiaban en el boca-a-boca tradicional, hoy en día el acceso a las redes sociales, donde se publican millones de opiniones y reseñas de otros usuarios sobre cierto producto o servicio, ha abierto un nuevo canal informativo para los usuarios, los cuales cada vez más tienden a consumir dicha información, pero también contribuyen a crearla. En concreto, en TripAdvisor se publicaron alrededor de 73 millones de reseñas y opiniones en 2022 (esto incluye 30,2 millones de reseñas y 31,6 millones de fotos y vídeos) (TripAdvisor Transparency Report, 2023).

En el apartado siguiente, se proporcionará una definición del término CGU, así como una enumeración de sus diferentes características. Además, se mostrarán los diversos tipos de CGU que existen hoy en día, haciendo hincapié en las reseñas generadas por los usuarios, que son el tipo de CGU en que se centra la presente Tesis.

### 2.1.2. Definición y características de CGU

En el siguiente apartado y retomando el tema del eWOM, dentro del cual se encuadra el CGU, se van a revisar las definiciones de eWOM y CGU que se ofrecen en la bibliografía consultada y se analizarán las características de este tipo de contenido digital. Las investigaciones sobre las reseñas de productos publicadas en plataformas en línea son relativamente nuevas, lo que se

puede observar por la falta de acuerdo en una denominación concreta para estas (Candel-Mora, 2015: 38). De esta forma, encontramos múltiples designaciones como «electronic word of mouth» o «eWOM» (Pollach, 2006), «online consumer reviews» (Vásquez, 2012) «user generated product reviews», «product reviews» or «user opinions» (Ricci y Wietsma, 2006).

Como se ha mencionado al inicio de este apartado, las denominaciones elegidas para la presente tesis serán «contenido generado por el usuario» (CGU) y «reseñas generadas por el usuario» (RGU), entendiéndose el CGU como un concepto más general que engloba la totalidad de contenidos generados por el usuario en Internet, dentro de los cuales hay diferentes tipos, entre ellos las reseñas generadas por los usuarios o RGU. En el apartado siguiente se definen los conceptos de eWOM y UCG para comprender su origen e importancia.

#### 2.1.2.1. Definición de eWOM y CGU

Son múltiples las definiciones que se han dado de CGU a lo largo del tiempo y, como se verá a continuación, este no se puede entender como un concepto aislado e independiente, sino como un concepto aparecido dentro de la gran red de Internet. De esta manera, algo que caracteriza al CGU es que este se puede crear, compartir o modificar de manera individual o colaborativa y que está muy relacionado con la comunicación boca a boca electrónica o *electronic word-of-mouth* (eWOM) (Smith, Fischer y Yongjian, 2012: 103).

Se podría definir el eWOM como cualquier forma de interacción o comentario positivo o negativo de un cliente potencial, actual o antiguo sobre una marca, producto o servicio, disponible para otros consumidores y organizaciones a través de un medio digital (Internet) (Hennig-Thurau, 2004; Tucker, 2011: 38; Vásquez, 2014: 66). Los comentarios se convierten entonces en activos comerciales muy valiosos (Amaral, Tiago y Tiago, 2014: 140; Smith, Fischer y Yongjian, 2012: 103). El CGU se podría definir como las contribuciones o aportaciones de contenidos no editados profesionalmente que cualquier usuario puede realizar de manera pública en Internet, a través de redes sociales o de plataformas en línea, y que no constituya un modo de vida o copia (Fernández-Castrillo, 2014: 60; García de Torres, 2010: 585; Hermida y Thurman, 2008; Hosteltur: 2018; Presi, Saridakis y Hartmans y 2014: 02). En la misma línea, Manap y Adzharudin (2013: 54) afirman lo siguiente:

User generated content also can be defined as the aggregation and leveraging of users' content on the Internet. It is related to the digital transformation of objects: written sounds, text, and images. Content is neither the graphical user interface nor the commands and text given by the administrators of the site. Rather, it is created

and uploaded by users of the platform, and users can be either organizations or individuals.

El CGU lo elaboran usuarios no profesionales que contribuyen voluntariamente con datos, información o archivos multimedia que aparecen en la Web para ayudar o entretener a otros usuarios (Fernández-Castrillo, 2014: 60; Krumm, Davies y Narayanaswami, 2008: 10). Blackshaw y Nazzaro (2006: 04) y Manap y Adzharudin (2013: 54) definen el CGU como los comentarios, perfiles y fotografías en línea hechos por los consumidores, y particularmente mencionan a los viajeros. Boyd y Ellison (2008), Smith, Fisher y Yongjian (2012: 102) aseguran que el CGU es un medio importante a través del cual los consumidores se pueden expresar y comunicarse con otras personas a través de Internet. Manap y Adzharudin (2013) se centran más en los aspectos que aporta el CGU y lo define como la acción de agregar valor a través de la creación de contenido digital y el aprovechamiento de dicho contenido por parte de los usuarios.

De las definiciones anteriores, se puede concluir que los autores coinciden en que el CGU es un medio de comunicación accesible, en que personas no profesionales pueden publicar su contenido y expresarse de manera pública a través de Internet para comunicarse con otros consumidores.

Según Pollach (2006: 03), el principal propósito del CGU es «to inform potential buyers of the strengths and weaknesses of consumer products», a lo que Vásquez (2012: 111) añade que el principal propósito de las reseñas de productos en línea es calificar, evaluar y describir, además de dar recomendaciones a otros usuarios a favor o en contra de un producto o servicio concreto.

Así, se aprecia que uno de los principales propósitos de los consumidores al escribir estos contenidos es el de compartir sus experiencias sobre un producto o servicio, así como el de evaluar las fortalezas y debilidades de dichos productos o servicios.

En conclusión, se podría decir que el CGU es un tipo de contenido realizado de manera pública en Internet por personas no profesionales, cuyo propósito es el de informar a otros consumidores y evaluar las experiencias vividas antes, durante y después de un viaje. Ya sean comentarios con contenidos positivos o negativos, los otros usuarios pueden participar mediante contestaciones a dichos contenidos, por lo que se podría afirmar que es un contenido colaborativo que puede ser utilizado no solo por otros consumidores, sino por compañías y organizaciones.

En las definiciones anteriores se ha hablado del CGU como un género que engloba diferentes tipologías de contenidos. No obstante, en una sociedad donde las personas viajan y se desplazan cada vez más, parece totalmente normal que, en gran parte de la bibliografía estudiada, los autores que tratan el CGU, hablen directamente de reseñas y, más concretamente, de reseñas en plataformas en línea de productos o servicios turísticos o de restauración. Además, el comportamiento de compra de los turistas ha cambiado gracias al desarrollo de nuevas tecnologías y a la aparición de diferentes plataformas y sitios web de reservas en línea, las cuales atraen el interés de los consumidores mediante las evaluaciones de servicios que les ayuden a encontrar lo que buscan para sus estancias turísticas (Chatzisavva, 2017:22; Comisión Europea, 2014).

Ricci & Wietsma (2006) se refieren al CGU en el sector del turismo y lo define como «the evaluation of users posted on a travel review site on their experience». Por otra parte, Ricci y Wietsma (2006: 297) describen las reseñas de productos como piezas subjetivas y no estructuradas de texto que describen los conocimientos, las experiencias y las opiniones del cliente sobre el producto junto con una valoración final de dicho producto. Por este motivo, en la presente tesis, el objeto de estudio principal serán las reseñas de productos o servicios turísticos y de restauración escritas por los usuarios en plataformas en Internet, las cuales se definirán y analizarán más adelante.

En el siguiente apartado se expondrán las razones por las que tener en cuenta el CGU y las RGU es primordial no solo por parte de los consumidores, los cuales buscan información sobre sus viajes, sino también por parte de los proveedores de este tipo de productos o servicios. Una buena estrategia y gestión de estos contenidos puede suponer una gran diferencia en la imagen de la empresa y en sus ventas. A continuación, se observará el porqué.

### 2.1.3. Importancia y estudio del CGU

Como se ha avanzado en apartados anteriores una buena gestión del contenido generado por los usuarios es vital para las empresas proveedoras de servicios y productos turísticos y de restauración. El CGU ayuda a dichos proveedores en sus campañas de marketing, no solo para anunciarse en redes, sino también para la captación de clientes y para la conversión de sus búsquedas en Internet en ventas. Los usuarios por su parte también sienten una mayor seguridad cuando compran algo que recomienda otro consumidor. De esta manera, se crean interesantes sinergias que benefician a todas las partes implicadas. A continuación, se detallan todos estos aspectos empezando por la importancia del CGU y mediante el estudio



multidisciplinar de este tipo de contenido en su vertiente empresarial y de marketing, desde el punto de vista lingüístico y entrando en los aspectos más psicológicos de las reseñas en Internet.

#### 2.1.4. Importancia del CGU

Según un artículo de la Harvard Business Review (Kelly, 2012), el 72,1 % de los consumidores pasa la mayor parte del tiempo en sitios web traducidos a su idioma. Además, el 72,4 % de dichos consumidores afirmó que es más probable que acaben comprando un producto si la información está en su lengua. Por último, la encuesta mostró que al 56,2 % de los encuestados no le importaría pagar un precio más alto por un producto si la información relativa a este está en su propia lengua. Esto lo corroboran Common Sense Advisory (2016), además de Görög y Sánchez-Gijón (2014), los cuales afirman que los usuarios están más dispuestos a comprar si la información sobre el producto o servicio está en su idioma. Es más, los consumidores están dispuestos a pagar más si el producto en sí está en su lengua.

Así, es posible afirmar que el CGU es importante a muchos niveles en los sectores del turismo y de la restauración. Más todavía si tenemos en cuenta que las compras de este tipo de productos y servicios se consideran de alto riesgo (Hidalgo-Alcázar, Sicilia y Ruiz de Maya, 2015: 77; Schemmann, 2011: 01). En palabras de Ricci y Wietsma, (2006: 02):

[...] reviews are expected to provide a more nuanced view, supported by ratings and rich media data (photos, video, etc.) [...] Products, or better said tourist services, lacks the feature of “trybefore-buy” or “return in case the quality is below expectance” (Buhalis, 2003). This implies that (online) tourist product buying involves a certain amount of risk taking, which could be lowered by providing the user a better and more reliable product description. Research has already proved that consumers put more trust in fellow consumers than in content provided by (marketing) agencies.

Los consumidores no pueden probar el servicio de antemano, ya que se trata de productos y/o servicios intangibles, por este motivo, recurren a Internet para encontrar reseñas o información útil de otros viajeros, es decir, CGU que pueda ayudarles en su decisión de compra y que minimice la sensación de riesgo e incertidumbre (Presi, Saridakis y Hartmans, 2014: 11; Schemmann, 2011: 01). Es más, los usuarios prefieren recopilar información a través de este tipo de redes sociales o buscadores que mediante los canales más tradicionales (Amaral, Tiago y Tiago, 2014: 138; Banerjee y Chua, 2014: 01).

Este tipo de contenido en línea ofrece, además, acceso a información real escrita por otras personas y que no ha sido suavizada o censurada por los medios informativos (Krumm, Davies y Narayanaswami, 2008: 10). Hoy en día, los consumidores confían más en este tipo de contenidos que en la publicidad o los contenidos originados para las campañas profesionales de marketing de las empresas. Hidalgo-Alcázar, Sicilia y Ruiz de Maya (2015: 80) afirman que «el boca-a-boca online juega un papel importante en la adquisición y la retención de los consumidores en la era del comercio electrónico». De esta manera, un cliente puede querer expresar su (in)satisfacción con un producto o servicio a través de *word-of-mouth* (WOM) durante cualquier momento del servicio, es decir, que puede leer y utilizar la información antes del servicio, durante este o incluso después (Amaral, Tiago y Tiago, 2014: 137; Presi, Saridakis y Hartmans, 2014: 02), ya que los usuarios quieren informarse sobre su destino antes, para ayudar a su decisión de compra; durante, para elegir lugares que visitar o restaurantes donde comer; y después de su viaje, para narrar su experiencia en el destino visitado y para intentar ayudar a otros viajeros.

Se debe tener en cuenta que las reseñas negativas tienen un mayor impacto sobre las ventas que las reseñas positivas y que estas disuadirán a los consumidores de adquirir determinado producto o servicio (Chevalier y Goolsbee: 2003; Mayzlin: 2006; Pollach, 2006: 2; Presi, Saridakis y Hartmans, 2014: 03; Purnawirawan et al., 2015; von Helversen et al., 2018: 01). Es más, los consumidores consideran que el eWOM negativo es más útil que el eWOM positivo para distinguir entre productos de alta calidad y productos de baja calidad (Pollach, 2006: 02). De este modo, aquellos productos con mejores valoraciones y, por tanto, eWOM positivo tienen unas mayores ventas, puesto que este influye en el criterio de los compradores, así como sus intenciones de compra de un determinado producto o servicio. El 92 % de los consumidores confía en las reseñas y opiniones de los otros internautas antes que en la publicidad y el 66 % de las búsquedas de información que se realizan antes de adquirir un producto o servicio turístico corresponde a la búsqueda de reseñas en línea (Hosteltur: 2018).

Chik y Vásquez (2016: 05), Dhar y Chang (2009: 301) y Pollach (2006: 02) también se muestran de acuerdo en la afirmación de que la mayoría de los usuarios prefieren recomendaciones de otros consumidores antes que las elaboradas por críticos profesionales o por empresas. Autores como Barreda y Bilgihan (2013: 267), Buhalis y Law (2008), Gretzel et al. (2007), Ricci y Wietsma (2006: 02), Tucker (2011: 37), Vásquez (2014: 69) o Zehrer, Crotts y Magnini (2011: 108), también constatan esta estadística añadiendo que la mayoría de los viajeros opina que los comentarios proporcionan información más actualizada, creíble y de mayor calidad que la ofrecida por otras fuentes de información.

Según un estudio de Destination Analysts (2019)<sup>6</sup>, las aplicaciones más utilizadas por los usuarios que querían planear un viaje al extranjero fueron tanto las de agencias de viajes digitales (Expedia, Hotels.com, etc.), como las plataformas de reseñas (TripAdvisor, Yelp, etc.). Sin embargo, en dicha encuesta se aprecia un descenso del 13 % en 2019 en el uso de dichas aplicaciones para la planificación de los viajes con respecto al año 2018. En este estudio también se afirma que las reseñas generadas por los usuarios todavía no han llegado a su punto álgido, y que el WOM tradicional cara a cara, sobre todo de familiares, sigue siendo el preferido porque los consumidores confían más en él. No obstante, no niegan el potencial de las reseñas, puesto que ofrecen información y opiniones de gran utilidad, y añaden que, si estas además se acompañan con vídeos o fotos, todavía se perciben como más creíbles.

Estos contenidos afectan a la imagen percibida del producto turístico, es decir, cómo valoran otros turistas el destino visitado y qué tal fue su experiencia (Hidalgo-Alcázar, Sicilia y Ruiz de Maya, 2015: 75; Smith, Fisher y Yongjian, 2012: 102). Por tanto, las opiniones de otros viajeros condicionan las ideas que los receptores del CGU se forman del producto turístico, es decir, la dimensión cognitiva de la imagen, si dichos receptores reconocen al autor del CGU como experimentado o de confianza (Hidalgo-Alcázar, Sicilia y Ruiz de Maya, 2015: 75; Litvin, Goldsmith y Pan: 2008).

Para adaptarse a estas nuevas corrientes digitales, el turismo ha incorporado la tecnología al sector, empezando por desarrollar sistemas informáticos de reserva y para interactuar con los clientes, así como herramientas de búsqueda y demás (Cooper y Hall: 2007). Es más, según Dhar y Chang (2009: 300), Internet ha posibilitado «a new era of user-generated content, threatening the hegemony of traditional content generators as the primary sources of “legitimate” information». Es decir, que los consumidores de CGU tienen acceso a opiniones de otros consumidores fuera de su círculo de conocidos por lo que dichas fuentes tradicionales pasan a un segundo plano.

De esta manera, entender las motivaciones de los consumidores al redactar cierto CGU tras tener una experiencia negativa se convierte en algo crucial (Presi, Saridakis y Hartmans, 2014: 05). Solo de esta manera las empresas de productos o servicios turísticos podrán utilizar dichas reseñas para apoyar sus campañas de marketing y sus ventas. Asimismo, las reseñas se

---

<sup>6</sup> Contenido extraído de: <https://www.destinationanalysts.com/welcome-to-2019-american-traveler-sentiment-weakens-2-2/>

convierten en un elemento de utilidad para los consumidores que buscan organizar su futuro viaje o para los que ya están de viaje, pero buscan recomendaciones.

En los apartados siguientes, se ha desglosado la investigación en CGU desde los puntos de vista desde los que este se ha estudiado hasta ahora, para tener una visión clara de sus ámbitos de aplicación y sus efectos. Esto permitirá observar los puntos de encuentro entre las diferentes disciplinas en cuanto a las características del CGU e identificar el impacto positivo de un CGU de calidad gestionado correctamente.

#### *2.1.4.1. Estudio multidisciplinar del CGU*

Pese a que el CGU es un concepto relativamente nuevo, este se ha examinado siempre desde diferentes perspectivas y ámbitos, como el empresarial y de marketing. También se han estudiado los aspectos más psicológicos del CGU, lo que ha servido para entender por qué los usuarios elaboran CGU y qué efecto tiene este sobre otros consumidores. Sin embargo, también es crucial entender los aspectos más lingüísticos de este, de esta manera, en el presente apartado se observarán las principales características del CGU desde estos tres puntos de vista.

Es evidente que el CGU resulta muy útil para los negocios que comercializan productos y servicios online, por lo que se han estudiado estos puntos de vista haciendo especial énfasis en el impacto del CGU para el sector turístico.

En primer lugar, se ha examinado el CGU desde el punto de vista empresarial y de marketing. Dada la importancia de las RGU, han aparecido numerosos estudios que examinan y evalúan la relación entre las RGU y las ventas de los productos. Por ejemplo, los realizados por Clemons, Gao y Hitt (2006); Dellarocas et al. (2004); Li y Hitt (2004) o Gao, Gu y Lin (2007). Gracias a estos estudios realizados desde el punto de vista del comercio y la empresa, se ha podido observar que en esencia el WOM les ayuda a saber lo que los potenciales clientes opinan sobre la marca (Williams et al., 2008: 117) por lo que el CGU y, en especial, las reseñas tienen gran importancia y pueden resultar muy beneficiosas. Por un lado, ayudan a que los turistas se puedan informar y facilitan las transacciones turísticas (Manap y Adzharudin, 2013: 55). Por otro lado, la influencia de las RGU tiene un gran alcance, por lo que se ha podido establecer una relación directa entre las RGU y la compra de productos o servicios (Vásquez, 2014: 69).

Es esencial disponer de un servicio electrónico de calidad y, para conseguirlo, los proveedores de productos o servicios turísticos deberían monitorizar el sitio web y abordar las publicaciones potencialmente problemáticas para evitar posibles dificultades (Barreda y Bilgihan, 2013: 268;

Presi, Saridakis y Hartmans, 2014: 03; Smith, Fisher y Yongjian, 2012: 111). Si se analizan las opiniones de los clientes, se llega a comprender mucho mejor la experiencia que han tenido en sus destinos. Así, se pueden responder de forma directa las quejas y mejorar aquellos aspectos con que los usuarios están descontentos (Pan, Maclaurin y Crofts: 2007). Por este motivo, se recomienda a los anunciantes que incentiven a los consumidores a participar en la creación de este tipo de contenidos de manera activa, además de validar y responder a los comentarios realizados por otros usuarios. También se aconseja a las empresas u organizaciones que creen campañas virales de marketing y de WOM (Barreda y Bilgihan, 2013: 268; Wells et al., 2000).

Pese a esta necesidad de manutención y monitorización, el CGU es una estrategia de marketing económica, ya que las empresas no tienen que pagar nada por estos contenidos, lo cual ahorra tiempo y tiene un alcance 10 veces superior en redes (Hosteltur: 2018; Mayzlin, 2006). Además, se trata de un contenido más auténtico y efectivo (Hosteltur: 2018). Según Hosteltur (2018), este tipo de contenido mejora la interacción y aumenta el tráfico, lo que resulta en un aumento de las ventas. También ayuda a diseminar la mala reputación digital de una marca (Dellarocas, 2003), a comprender mejor las reacciones de los consumidores con respecto a una nueva línea de productos o servicios y a descubrir nichos o áreas que estén teniendo un peor desempeño (Barreda y Bilgihan, 2013: 269).

Las principales ventajas que el CGU puede aportar a una campaña de marketing en el sector del turismo son las siguientes (Amaral, Tiago y Tiago, 2014: 139):

- a. Mejora la percepción del producto, servicio o marca y ayuda a gestionar la reputación de estos.
- b. Ayuda a fidelizar a la clientela.
- c. Facilita las investigaciones de mercado, la generación de una idea de la marca por parte del cliente y el lanzamiento de servicios o productos.
- d. Amplía las posibilidades del marketing WOM.
- e. Facilita la comunicación con los empleados y las relaciones públicas.

Los diferentes autores que han estudiado el CGU desde este punto de vista empresarial están de acuerdo en que este ayuda a la creación de la marca aumenta y mejora la visibilidad del producto o servicio y ayuda a los consumidores a tomar decisiones de compra tras informarse, lo que conlleva un aumento de las ventas. Además, puesto que los consumidores ya sabrán lo que están comprando exactamente, hay un menor riesgo de que devuelvan el producto o pidan un reembolso. No obstante, para sacar partido a todas estas ventajas, hay que prestar mucha

atención y mantener de manera continuada las plataformas de reseñas donde se publican. Por ejemplo, en plataformas como TripAdvisor, las reseñas requieren gran atención dada su gran visibilidad y su relevancia para determinados proveedores de servicios (Presi, Saridaki y, Hartmans, 2014: 19).

TripAdvisor se presenta como una plataforma web de viajes con millones de reseñas y opiniones de viajes hechas por turistas de todo el mundo. En julio de 2017, el número de usuarios de TripAdvisor llegó a un máximo histórico de 324 millones (TripAdvisor, 2017: 09). Es más, en 2022, el total de reseñas sobre hoteles fue de más de 11 millones y el de restaurantes de más de 10 millones (TripAdvisor Transparency Report, 2023). Para Munar et al. (2013), TripAdvisor promueve la interacción social en la industria del turismo, animando a los usuarios a compartir sus experiencias en las distintas áreas de los sectores de la hostelería y el turismo.

En el caso de este tipo de plataformas, Kreck (1998) y Pullman et al. (2005) aconsejan analizar los comentarios de los consumidores para comprender las experiencias que estos han tenido con los productos o servicios. Gracias a este análisis, será posible mejorar aquellas áreas con las que los clientes están insatisfechos para no volver a tener reseñas negativas que puedan dañar la imagen de la marca. Solo así se conseguirá un mejor servicio, unos mayores beneficios y una mejor imagen. Del mismo modo, Barreda y Bilgihan (2013: 267) afirman lo siguiente:

[...] the most appropriate way to gain a deep comprehension of a costumers' feeling about their experiences with the hotel is analyzing the content of the customer's comments. This analysis of reviews has been largely utilized to improve areas in service and allocate resources in needed operational areas in effective manners (Kreck, 1998). Hotel operators and managers understand that information provided in a form of reviews or ratings enables them to make management decisions that ended in improved service operations, maximized profits, optimized revenues, and increased of brand loyalty and brand image (Ford and Bach, 1997).

De este modo, se podría afirmar que, si los proveedores de servicios o productos turísticos desarrollan una estrategia de marketing en sus redes sociales y en las plataformas de reseñas en las que participen, lograrán conseguir una mayor interacción y CGU de sus clientes, el cual, si se gestiona correctamente será muy beneficioso. No obstante, aunque las reseñas puedan ser una gran ayuda tanto para negocios como para consumidores, también tienen sus inconvenientes puesto que muchas empresas llevan a cabo campañas de marketing desleales, en las que crean contenido falso o que tiene como objetivo dañar la imagen de sus competidores.

De esta manera Banerjee y Chua (2014: 01) afirman que:

[...] users' growing penchant for reviews has resulted in the rise of deceptive opinion spamming, which involves posting misleading reviews to influence users' impression on products and services insidiously. As a result, it has become challenging for users to distinguish between genuine and deceptive reviews.

Es decir, que, con el auge de las redes sociales y el constante crecimiento del número de reseñas en línea, aparecen de manera cada vez más frecuente reseñas engañosas o falsas, cuyo propósito principal es el de influenciar a los usuarios para que no compren determinado producto o servicio o para que adquieran otro de la competencia. Según Valant (2015: 02) en un informe del Parlamento Europeo, las reseñas falsas representan entre el 1 % y el 16 % de todas las reseñas escritas por consumidores en Internet. Wu et al. (2010) definen las reseñas falsas o *shill reviews* como reseñas que «distort popularity rankings given that the objective is to improve the online reputation» (Ong, Mannino y Gregg, 2014: 69), es decir, reseñas que influyen en la reputación de un producto o servicio de manera que distorsionan su popularidad entre los consumidores y merman su reputación en Internet.

Normalmente los creadores de estas reseñas falsas o engañosas suelen ser compañías que intentan manipular información mediante la publicación de falsas reseñas positivas de sus propios productos o de reseñas negativas sobre los productos de sus competidores (Chatzisavva, 2017: 25; Lappas et al., 2016). Banerjee y Chua (2014: 01) llaman a esta nueva tendencia y mala práctica en Internet *opinion spamming*. Estos autores van un paso más allá y distinguen dos tipos distintos de reseñas maliciosas. Por un lado, las reseñas engañosas o *deceptive* y las reseñas disruptivas o *disruptive*. Las primeras son aquellas que contienen contenido totalmente irrelevante. Las segundas son aquellas que se escriben de manera consciente para que parezcan reales, por lo que no son fáciles de detectar (Banerjee y Chua, 2014: 01). Chatzisavva (2017: 25) distingue los siguientes tipos de reseñas engañosas:

[...] (1) objectively inaccurate evaluations; (2) evaluations that are not genuine but have been drawn up by consumers or other parties for the purpose of deception; and (3) misleading advertising and unfair marketing practices by supervisors or others.

Es decir, que entre los distintos tipos de reseñas se pueden encontrar las que son imprecisas y tal vez no cuadran con el producto que se anuncia, otras que han sido escritas por terceras personas simplemente para engañar a los consumidores y otras que son publicidad engañosa y que han sido escritas por una empresa para dañar la imagen de su competencia.

Para distinguir entre una reseña real y una engañosa, Banerjee y Chua (2014: 02) identifican tres rasgos identificativos que nos permiten diferenciarlas: el género, el estilo y la legibilidad de la reseña. Con legibilidad se refieren al esfuerzo de comprensión y la experiencia que debe tener un usuario cuando lee determinada reseña. Para escribir una reseña falsa, será necesario un mayor esfuerzo cognitivo que para una genuina, lo que influye de manera directa en la forma que los lectores percibirán dichas reseñas (Banerjee y Chua, 2014: 02). En cuanto al género, Banerjee y Chua (2014: 02) distinguen dos tipos de géneros textuales, el informativo (reseñas reales) y el imaginativo (reseñas falsas). Por último, en cuanto al estilo, afirman que en las reseñas genuinas se encuentran muchas más referencias a experiencias propias (primera persona) y verbos en pasado que en las reseñas falsas.

Sabiendo esto, los usuarios y lectores que quieran informarse a través de reseñas antes de adquirir cierto producto o servicio podrán analizar estos textos para asegurarse de que lo que leen son comentarios genuinos de compradores reales. Desafortunadamente, cada vez es más complicado distinguir reseñas genuinas de reseñas engañosas y, en ocasiones, por culpa de estas malas prácticas la confianza de los consumidores se ve perjudicada.

En segundo lugar, se ha analizado el CGU desde el punto de vista cognitivo, ya que para poder gestionar el CGU en sus plataformas, los proveedores de productos y servicios turísticos deben entender los motivos por los que los usuarios escriben sus reseñas. A esto se suma también la importancia de localizar y gestionar las reseñas falsas, cuya intención es dañar la imagen de los productos y servicios en cuestión. A continuación, se han identificado las razones para la elaboración de dichos contenidos y los tipos de reseñas engañosas identificadas en los diferentes estudios sobre CGU y reseñas en Internet.

Cuando se habla de la imagen que un producto tiene de cara al cliente, se hace referencia a aquello «que los turistas piensan, sienten y asocian con relación a un determinado producto turístico» (Hidalgo-Alcázar, Sicilia y Ruiz de Maya, 2015: 76). Esta imagen se conforma por dos componentes, el componente cognitivo y el afectivo. Por un lado, el componente cognitivo está formado por las creencias, los conocimientos y las evaluaciones que el turista tiene sobre las características del producto turístico. Por otro lado, el componente afectivo está ligado a las motivaciones y los deseos en la toma de decisiones del usuario, así como a los sentimientos que los consumidores tienen con respecto al producto turístico (Gartner: 1993; Hidalgo-Alcázar, Sicilia y Ruiz de Maya, 2015: 76; Pike y Ryan: 2004).



También en relación con los sentimientos de los consumidores a la hora de escribir este tipo de contenidos, Dhar y Chang (2009: 301) afirman que los clientes escriben sus reseñas por diferentes razones que pueden ser las siguientes:

- a. Mejorar su influencia o estatus.
- b. Crear un contenido que pueda ser de ayuda a otros usuarios.

Presi, Saridakis y Hartmans (2014) identifican además como emociones las intenciones altruistas, la venganza, la intención de desahogarse de sentimientos negativos, el realizarse a uno mismo, las motivaciones económicas y la extraversión. Según Krumm, Davies y Narayanaswami (2008: 10):

For content suppliers, the process can be rewarding because it lets them receive recognition for their contributions. For consumers, besides the potential to inform or entertain, the content gives a glimpse into real data from other people, unsanitized by regular media outlets.

Es decir, la elaboración de estas reseñas puede resultar muy satisfactoria tanto para los consumidores, que son también escritores de dicho contenido, como para los proveedores de productos y servicios. Los primeros no solo obtienen información de otros usuarios sobre un producto o servicio, sino que también obtienen un reconocimiento por su contribución. Los segundos reciben reconocimiento por sus esfuerzos y también una valiosa herramienta para identificar puntos débiles en su oferta de productos y servicios.

Por último, se ha estudiado el CGU desde el punto de vista lingüístico, ya que uno de los objetivos de la presente tesis ha sido analizar la importancia de una correcta gestión lingüística del contenido, empezando por los daños que puede causar una traducción automática sin una adecuada estrategia de posesición que tenga en cuenta los principios de la localización.

Al igual que en el mundo *offline*, en los sitios web y en las plataformas de reseñas de Internet, los consumidores siempre han querido narrar y compartir sus experiencias con otros consumidores sin que nadie les haya instado a hacerlo, es decir, por su propia voluntad (Pollach, 2006: 01). Estos narran sus historias utilizando diversos formatos y con unas características determinadas, las cuales se analizan a continuación.

Vásquez (2012: 109) identifica la narrativa en orden cronológico como una de las principales características de las RGU en el sector del turismo. El turista narra su experiencia temporal en un hotel de manera lineal en ocho fases, las cuales van desde la fase inicial de la planificación y

la primera vez que se entra a la habitación, hasta el registro de salida y las comunicaciones con el hotel posteriores a la visita. Estas se enumeran en la Tabla 2:

Tabla 3. Fases de la experiencia temporal de un cliente (Vásquez, 2012: 109).

1. Fase de investigación y planificación: incluye la lectura de reseñas en plataformas de Internet.
2. Fase de reserva de una habitación: se puede hacer por Internet, mediante una agencia de viajes o con el móvil.
3. Primeras impresiones del hotel y el personal, puede darse en el hotel mismo o incluso antes de llegar a él mediante las comunicaciones previas por correo electrónico.
4. Realización del <i>check-in</i> .
5. Primeras impresiones de la habitación.
6. Actividades relacionadas con la estancia. Como el trato con el personal, el restaurante o los servicios e instalaciones del hotel.
7. Realización del <i>check-out</i> .
8. Comunicación con el hotel después de la estancia.

Estas narrativas tienen una estructura determinada que incluye un resumen, antecedentes (razón del viaje, acompañantes, etc.), valoración explícita, interacción con el personal del hotel, final de la estancia (*check-out* y costes) y consejos o sugerencias personales (Vásquez: 2011), aunque no todas cuentan con todas las partes mencionadas anteriormente. Otras características asociadas con el discurso según Candel-Mora (2015: 39) y Vásquez (2012) son:

- El discurso indirecto.
- Pequeñas introducciones o prefacios a la historia de la experiencia.
- Cambios deícticos, los cuales conectan al autor con los lectores del contenido.
- El uso del humor, el detalle y la experiencia personal.

Además, en el discurso escrito no hay posibilidad de utilizar rasgos no lingüísticos, como pueden ser signos o gestos, los cuales pueden ser cruciales a la hora de interpretar lo que se quiere decir,

dificultad que Vásquez (2012: 107) tiene muy en cuenta a la hora de realizar este tipo de investigación.

Según Pollach (2006: 08) otras características específicas de este género son, por ejemplo, la intertextualidad o referencias a comentarios anteriores, el perfil personal del reseñador y otros elementos paralingüísticos, como estrategias ortográficas utilizadas para compensar la impersonalidad del discurso escrito (mayúsculas, signos de exclamación o de interrogación, signos de puntuación, emoticonos y acrónimos, entre otros). Pollach (2006: 03) afirma también que estos textos están aislados de conversaciones previas entre consumidores, es decir, los consumidores escriben nuevas reseñas o comentarios que no tienen por qué seguir el hilo de reseñas o comentarios anteriores, si no que cada autor escribe su reseña sobre su propia experiencia personal independientemente de lo que otros autores hayan escrito anteriormente (Mackiewicz, 2008; Pollach, 2006: 03).

Este rasgo distingue este tipo de texto de otros géneros y tipos de interacción entre consumidores. Además, Pollach (2006) menciona otros elementos como la persistencia, la asincronicidad y la potencial ausencia de respuestas o *feedback*. Esto significa que los autores de las reseñas crean textos aislados e independientes de las reseñas escritas con anterioridad a las suyas, aunque esto no implica que los lectores no puedan iniciar una «conversación» a través de respuestas a las reseñas.

Por el contrario, Vásquez (2015) cree que la intertextualidad y la interdiscursividad son dos rasgos muy comunes en las reseñas en línea. Según Vásquez (2015), el número de reseñas en Internet ha crecido en los últimos tiempos y las prácticas discursivas asociadas con este género han cambiado y evolucionado, por lo que la autora asegura haber encontrado «numerous examples of intertextuality» (Vásquez, 2015: 68). Vásquez (2015) identifica diversos mecanismos en que esta intertextualidad se manifiesta en las reseñas:

1. Alusiones intertextuales a reseñas anteriores. En este caso podrán expresar su acuerdo o desacuerdo con estas reseñas anteriores.
2. Referencias intertextuales a otros tipos de textos relevantes para el producto. Por ejemplo, enlaces a Wikipedia, videos de YouTube y otros sitios web. En el caso concreto de las reseñas de productos o servicios turísticos menciona la referencia a páginas web de hoteles o a servicios de reserva en línea.
3. Referencia a fuentes y recursos *offline*. Se integran citas a otras fuentes textuales en las reseñas para proporcionar información adicional o aclarar ciertos aspectos del producto o servicio.

4. Alusiones o menciones explícitas de textos o referencias de cultura popular.

Además de incorporar esta clase de recursos intertextuales, las reseñas también pueden mezclar rasgos discursivos propios de otros géneros como, por ejemplo, textos para marketing, ventas o anuncios (Vásquez, 2015: 74). Algunos de estos rasgos discursivos se utilizan para que los autores de las reseñas se puedan dirigir de manera más directa a sus lectores más cercanos. De esta manera, Vásquez (2015) identifica los elementos sugeridos en la Ilustración 4:

Discourse features	Examples	Functions
Discourse markers	<i>Well, You know, Now, Oh</i>	Mark interactivity; mark text as "speech like"
Imperatives and second-person pronouns	<i>Don't plan on sleeping. Upon closer inspection you will see.</i>	Address readers directly
Interactive metadiscourse	<i>What can I say Let me tell you</i>	Appeal for readers' continued interest and participation in reading the text; mark text as "speech like"
Questions	<i>Can you believe that? would we stay here again?</i>	Simulate give-and-take of conversation; elicit some type of overall evaluation of experience
Answers	<i>and yes, I always use an oven thermometer</i>	Simulate give-and-take of conversation; demonstrate a reviewer's credibility/ethos
Questions and answers (simulated dialogue)	<i>Would you toss whole strawberries [...] into your \$40 blender? No. so ya think someones gonna die? Yup.</i>	Simulate give-and-take of conversation; alignment, shared perspective

Ilustración 4. Rasgos discursivos asociados a estrategias de acercamiento al lector (Vásquez, 2015: 35).

Los marcadores discursivos hacen referencia a expresiones como interjecciones o muletillas que añaden un tono de oralidad en el mensaje escrito. Vásquez (2015: 35) también menciona el uso de imperativos y pronombres en segunda persona, los cuales dan una sensación de cercanía, al igual que los marcadores metadiscursivos y las preguntas retóricas que además instan al lector a participar. En ocasiones, los escritores utilizarán también el recurso de las respuestas a preguntas que ellos mismos han formulado o de los diálogos simulados para dar credibilidad y veracidad a sus reseñas.

Por último, cabe destacar que existe la posibilidad de que los autores de las RGU no sean hablantes nativos de la lengua en que escriben (Pollach, 2006: 03), lo que puede dar lugar a formulaciones incorrectas o no propias de ese idioma, las cuales los autores importan desde sus lenguas nativas.

En resumen, se concluye que algunas de las principales características de las RGU serían las siguientes:

- Estructura: suelen tener una estructura similar a la de las historias y cuentan con introducción y narración de los hechos que, a menudo termina con el desenlace de la experiencia.
- Lenguaje oral: uso, en ocasiones excesivo e innecesario, de signos ortográficos como signos de exclamación o interrogación u otras estrategias de puntuación ortográfica. Uso de marcadores metadiscursivos que dan sensación de oralidad.
- Uso del humor y cambios deícticos.
- Narración de la experiencia personal: humaniza al escritor de la reseña y lo acerca al lector.
- Intertextualidad y referencia a comentarios anteriores: los lectores pueden contestar a las reseñas.

Como se puede observar, las RGU se han estudiado también desde el ámbito lingüístico y se han identificado sus rasgos de intertextualidad, sus características del lenguaje oral, su estructura, etc. No obstante, se ha apreciado la escasa literatura sobre las consecuencias de traducir estos contenidos utilizando sistemas de TA. Por este motivo, en la presente tesis, se retomará el estudio desde el punto de vista lingüístico y se aportará una visión novedosa sobre el uso de la TA para traducir reseñas del ámbito turístico y de restauración, la localización y la posesición de estas traducciones.

#### 2.1.5. Tipos de CGU: las RGU como nuevo género textual

Cuando se habla de CGU no se hace referencia solo a un tipo de contenido concreto, sino que existen diferentes clases de sitios web y plataformas en línea, dentro de las cuales se pueden identificar diversos tipos de contribuciones de usuarios. En muchos casos, incluso en un mismo sitio web o plataforma se encuentran diferentes tipos de CGU. En el presente apartado, se mostrarán algunos de estos tipos de contenidos, sobre todo las RGU de servicios y productos turísticos y de restauración. Se profundizará en las características principales de las RGU, en su influencia en las adquisiciones de los consumidores y en otros aspectos clave para entender este tipo de CGU.

En cuanto a los formatos de eWOM elegidos por los consumidores para escribir sus contenidos, se identifican los siguientes: reseñas de productos, hilos de conversación (foros), salas de chat en línea o *chatrooms*, video mensajes y sitios de quejas donde los consumidores pueden dialogar

e interactuar entre ellos (Barreda y Bilgihan, 2013: 266; Pollach, 2006: 03; Puri, 2007). En el caso de las plataformas en línea de servicios turísticos, se observa que TripAdvisor.com contiene foros, *chartooms*, reseñas, valoraciones, ideas y experiencias de viajeros que ayudan a otros viajeros a tomar decisiones en cuanto a sus destinos turísticos (Barreda y Bilgihan, 2013: 268). Otros intermediarios como Expedia, Orbitz o Travelocity también ofrecen la posibilidad de escribir comentarios y reseñas cortas de hoteles y de dar valoraciones numéricas en Internet (Barreda y Bilgihan, 2013: 265; Stringam y Gerdes Jr, 2010).

Los espacios interactivos como blogs y redes sociales (como Facebook, Instagram o Twitter) han posibilitado nuevas formas de interacción social, distintas de las conversaciones cara a cara o de los textos escritos tradicionales. Además, gracias a ellas, los consumidores están mejor informados que nunca y pueden interactuar y participar de maneras distintas gracias a la publicación de comentarios o valoraciones y de la edición de wikis (Banerjee y Chua, 2014: 01; Jones, Chik y Hafner, 2015: 01; Manap y Adzharudin, 2013: 53).

En sus inicios, la investigación en el campo de la comunicación en entornos digitales se centró en aquellos entornos que permitían interactuar emulando el lenguaje oral de manera síncrona (*chatrooms*) como de manera asíncrona (foros). También se han estudiado otros tipos de textos multimodales como los que aparecen en los blogs y en las redes sociales (Boyd y Heer 2006; Herring et al. 2005). No obstante, en los últimos tiempos el tipo de CGU que más paso se está abriendo y que más útil resulta a los consumidores son las RGU. Por esa razón, a continuación, se describirá el origen, las características y demás aspectos relevantes de este tipo de CGU, para crear conciencia de la repercusión y la influencia que estas pueden llegar a tener en los consumidores.

#### *2.1.5.1. Reseñas generadas por los usuarios (RGU)*

Como se ha indicado anteriormente, existen muchos de tipos de CGU, los cuales se han estudiado anteriormente, para comprobar el peso que tienen sobre la toma de decisiones de los usuarios a la hora de comprar un producto o servicio en Internet. Entre estos tipos de CGU, se encuentran las reseñas generadas por los usuarios de productos o servicios turísticos y de restauración publicadas en plataformas en línea, las cuales serán el principal objeto de estudio de la presente tesis. Estas tienen una especial transcendencia, puesto que disminuyen la percepción del riesgo y ayudan a los clientes a tomar decisiones informadas. A continuación, se exponen con detalle los tipos de RGU existentes y sus principales atributos.

Aunque se haya estudiado desde muchas perspectivas el impacto que tienen las RGU sobre el comportamiento de los consumidores, no se ha prestado mucha atención a los diferentes tipos de reseñas que se pueden encontrar en Internet hoy en día (Gretzel et al. 2007: 08; Schemmann, 2011: 02). Centrándose concretamente en las reseñas de productos y servicios turísticos, Schemmann (2011) identifica 12 tipos de reseñas clasificados en tres categorías que pueden variar dependiendo de la plataforma donde se escriban:

1. Valoración de servicios.
  - a. Textos estructurados y de escritura libre.
  - b. Valoraciones.
  - c. Recomendaciones.
  - d. Fotos y vídeos.
  - e. Resúmenes de reseñas y análisis de tendencias.
2. Para dar *feedback* y con función interactiva.
  - a. Valoración de la utilidad de la reseña.
  - b. Función de anotación o de respuesta a las reseñas.
  - c. Foros y comunidades.
3. Funciones para incrementar la utilidad a las reseñas.
  - a. Características de los reseñadores: se ofrece información sobre los escritores de las reseñas. Esto puede aportar, pero también restar credibilidad a dichas reseñas.
  - b. Uso del grafo social: por ejemplo, enlazando Facebook con TripAdvisor aprovechando las conexiones y contactos de esta red social.
  - c. Integración de las valoraciones en la búsqueda: por ejemplo, poder realizar la búsqueda de un hotel basándonos en las valoraciones de este.
  - d. Integración de reseñas y valoraciones de otras plataformas en una misma plataforma.

Según von Helversen et al. (2018: 02) las plataformas de RGU de normal ofrecen reseñas en dos formatos distintos:

1. Información estadística: puntuaciones que dan una visión general de la calidad de un producto.
2. Reseñas que contienen narraciones de las experiencias de los consumidores con un determinado producto.

A continuación, se han estudiado y resumido las diferentes definiciones de RGU que se han ofrecido desde el ámbito de la investigación. Se podría afirmar que las reseñas en línea son un género relativamente nuevo de CGU y sin un precedente análogo concreto (Vásquez, 2015: 77), puesto que estas empezaron a existir cuando aparecieron los primeros sitios webs y plataformas de opiniones en Internet. Anteriormente, si un consumidor deseaba obtener una opinión de otro consumidor, este la recibía mediante WOM tradicional oral. De la misma forma, si quería expresar su opinión sobre un producto o servicio, debía escribir una carta a la compañía proveedora. Es decir, los consumidores no tenían la posibilidad de leer reseñas de otros usuarios en Internet. De acuerdo con esta afirmación, Pollach (2006: 02) dice lo siguiente:

Previously, people shared such information only orally in the form of WOM communication with other consumers or wrote letters to companies, but did not have the opportunity to share their opinions with other consumers in a structured, written format. Also, consumers were not able to obtain product-related information from strangers.

Las RGU, también conocidas como eWOM, aparecieron como un nuevo género mediático que ya no estaba restringido a un grupo de críticos profesionales. Ahora las reseñas pueden ser escritas por cualquiera con acceso a Internet (Chik y Vásquez, 2016: 04-05; Hennig-Thurau et al., 2004). Una cantidad significativa del CGU que encontramos en Internet se da en forma de reseñas (Chik y Vásquez, 2016: 04; Egbert y Biber, 2013). Chik y Vásquez (2016: 04) dan cuenta de la repercusión de las RGU e indican lo siguiente:

A significant amount of the user-generated content that is found online appears in the form of some type of review. A recent estimate places the genre of online reviews as comprising 2–4 per cent of the searchable Internet.

Hoy en día, existen multitud de plataformas en línea como TripAdvisor.com, con foros, *chatrooms*, reseñas, valoraciones numéricas e ideas y experiencias de los consumidores, para ayudar a otros usuarios a concretar ideas y elegir sus destinos de viaje, alojamientos y demás. Estas reseñas se podrían definir como un tipo de CGU con evaluaciones que se publican en sitios webs o plataformas de terceros (Mudambi y Schuff, 2010: 186). Gao, Gu y Lin (2007: 01) las consideran como una extensión del WOM, en la cual los consumidores se influyen unos a otros mediante la diseminación de sus opiniones. Estas RGU suelen ser descripciones de productos o servicios que pueden ir acompañadas de fotos o enlaces a otros sitios web o a vídeos.



Estas son formas de comunicación mediada por ordenador asíncronas e interpersonales, en que los usuarios que las leen y los que las escriben no están conectados personalmente, es decir, estos no se conocen entre sí (Vásquez, 2015: 67). Precisamente, se diferencian de otros tipos de redes sociales en que los lazos que unen a los agentes que participan en ellas son más débiles que en las redes sociales tradicionales (Vásquez, 2015: 21-22). Las RGU son un género predominantemente escrito y se pueden considerar como un tipo de medio de comunicación social, puesto que poseen algunas características típicas de la Web 2.0 como, por ejemplo, el que sean participativas, colaborativas, generadas por los usuarios, dinámicas y ricas en información (Vásquez, 2015: 21).

Según Pollach (2006: 02) los compradores escriben las reseñas para relatar sus experiencias, buenas o malas, con los productos o servicios que consumen, para animar o prevenir a otros potenciales consumidores de comprar determinado producto. Por otra parte, Vásquez (2014) afirma que «while the primary purpose of online reviews is for consumers to evaluate a product or service, reviews can be considered hybrid texts, in that they often also include description and narration». Es decir, que las reseñas no son solo textos que evalúan numéricamente un producto o servicio, sino que los usuarios que las escriben suelen hacerlo incluyendo descripciones y sus experiencias en forma de narración.

A modo de conclusión, se podría afirmar que el CGU está cambiando la estructura de la información sobre destinos turísticos, su accesibilidad y el conocimiento y la percepción de muchos de los productos y los servicios turísticos existentes (Manap y Adzharudin, 2013: 54). Esto se observa en la manera en que las personas interactúan con los contenidos en Internet. Ya no solo los leen y los consumen, sino que se han convertido en creadores y en críticos no profesionales. Muchos autores como Manap y Adzharudin (2013: 55) afirman que las redes sociales han llegado para quedarse y que su impacto ha dado forma a la manera en que las empresas y los consumidores se comunican e interactúan con otros usuarios

Así, se podría decir que la tarea de la gestión del CGU y las RGU es crucial a dos niveles. Por un lado, para las empresas, ya que una correcta gestión de dichos contenidos influye de manera directa en las ventas. Es decir, así como las reseñas y el eWOM negativos pueden provocar un descenso en las ventas, las reseñas y el eWOM positivos pueden hacer que las ventas se disparen. Por otro lado, estos contenidos ayudan a los usuarios durante el proceso de adquisición de los productos o servicios. No solo les proporcionan información útil para que puedan llegar a conocer el producto antes de adquirirlo, sino que también ayudan a disminuir el sentimiento de inseguridad y de riesgo que provoca la compra de bienes o servicios

intangibles que no se pueden probar con anterioridad, como es el caso de los productos y servicios turísticos y de restauración.

De esta manera, el análisis de las reseñas traducidas mediante TA que se realizará en esta tesis demostrará que la correcta implementación de una estrategia de posesición que siga los principios de la localización es crucial para la gestión del CGU, concretamente de las reseñas generadas por los usuarios en plataformas de Internet. Además, servirá para recalcar la importancia de localizar el contenido que se elabora para publicarse en línea.

## 2.2. La Localización

Para entender mejor la disciplina que llamamos localización, se establecen los orígenes y la evolución de la localización, a partir de los estudios publicados tanto por profesionales como por investigadores en localización. En el presente apartado se resumen los orígenes, la definición y las características de la localización, así como las diferencias que esta presenta con respecto a la traducción.

### 2.2.1. Orígenes

La aparición de la localización se remonta aproximadamente a los años 80, por lo que se considera que es una industria relativamente nueva, si se compara con la larga tradición de otras disciplinas como la traducción. La nueva necesidad de localización aparece gracias al uso creciente de los ordenadores personales. De la misma manera, el *software* de estos ordenadores debía adaptarse a las necesidades de cada *locale* y a las demandas de los usuarios, que pedían que el *software* fuera más funcional y eficiente. Además, también creció la necesidad de que estos programas estuvieran traducidos de forma que fueran entendibles y fáciles de usar (Esselink, 2003: 22; Fernández-Costales: 2010: 37; García-Nevarés, 2016: 65-67; Pym, 2012: 190).

Este proceso se llevó a cabo de diferentes maneras según la empresa distribuidora de estos productos. Citando a Esselink (2003: 22):

The international expansion of software and hardware developers automatically triggered the need to localize the products for international markets. Initially, software vendors dealt with this new challenge in many different ways. Some established in-house teams of translators and language engineers to build international support into their products. Others simply charged their international offices or distributors with the task of localizing the products. In both cases, the

localization effort remained separated from the development of the original products.

Esselink señala que es en este momento, cuando los creadores y distribuidores de *software* se dan cuenta de estas necesidades y empiezan a internacionalizar y localizar sus productos. Es por esto, que en las empresas que comercializaban estos programas se comienzan a crear departamentos internos de traducción (*in-house translation*) y a contratar traductores autónomos externos a la compañía (Esselink, 2003: 22-23).

Más adelante, en los años 90, los productos multimedia y la documentación que los acompaña se comienzan a distribuir a escala mundial, gracias a la aparición y al desarrollo de Internet (Esselink, 2000: 05). A raíz de estos acontecimientos, se crean algunas de las principales asociaciones sobre localización. Es el ejemplo de la Localisation Industry Standards Association (LISA), cuya intención fue aunar esfuerzos entre localizadores y desarrolladores (Esselink, 2000: 06; Fernández-Costales, 2010: 39), el Localisation Research Centre (LRC) y la Globalization and Localization Association (GALA). No obstante, LISA desapareció en 2011. Gracias a la creación de estas asociaciones y a la aparición de nuevos proveedores de servicios lingüísticos, se consolida la «industria de la localización» como tal, en que, tradicionalmente, se han podido diferenciar dos sectores principales: el sector de los fabricantes y distribuidores de *software* y *hardware* y el sector lingüístico, es decir, el de los localizadores (Esselink, 2000: 07; Jiménez-Crespo, 2009: 02; Quah, 2006: 19).

El proceso de localización de productos multimedia sufrió cambios relevantes, como, por ejemplo, que anteriormente, estos productos se localizaban cuando estaban finalizados, lo que provocaba que su lanzamiento en otros países se retrasara. Más tarde, esto cambia y los productos se plantean siguiendo el concepto del *simship* o *simultaneous shipment*, lo que significa que se empezaron a comercializar todas las versiones de un producto de manera simultánea. Por este motivo, surgió la necesidad de utilizar herramientas de localización que aceleraran el proceso sin perjudicar la calidad (Esselink, 2000: 477).

Otro gran cambio fue, por ejemplo, el perfil de los traductores, que ahora deberían tener conocimientos informáticos y habilidades más técnicas, además de un muy buen conocimiento lingüístico (Arevalillo-Doval, 2000: 05; Esselink, 2002: 01). Por otro lado, Internet empezó a adquirir una gran relevancia en los procesos de globalización de los negocios, lo que repercutió en la localización de los sitios web, que tuvieron que adaptarse a las lenguas y las convenciones culturales de los diferentes *locales* (O'Hagan y Ashworth, 2002: 67).

A partir de los años 90, los sitios web que en un inicio solo se traducían desde el inglés a otras lenguas, se empiezan a traducir desde muchas otras lenguas al inglés (Jiménez-Crespo, 2008: 29; Pérez-Fernández, 2010: 53). Este fenómeno tan significativo se dio porque los mercados crecieron y los programas se volvieron más complejos, así la traducción desde una lengua hacia otra única lengua empezó a ser insuficiente (Pym, 2010: 190). Además, gracias a la aparición y localización de nuevos productos informáticos y contenidos en línea, los perfiles de los usuarios también cambiaron y se hizo todavía más patente la necesidad de localización de los productos para que estos fueran mucho más accesibles (Fernández-Costales, 2010: 37; Pérez-Fernández, 2010: 54).

En los años 90, aparecen también nuevos conceptos en materia de localización, como la denominada «localización parcial», la cual consistía en la traducción no solo de la caja que contiene el producto, sino también en la localización de la documentación que lo acompaña, la interfaz y los subtítulos en caso de haberlos. Otro concepto es el de «localización total», mediante el cual se traduce todo lo mencionado anteriormente y se doblan las conversaciones y escenas (en el caso de los videojuegos) (Bernal-Merino, 2011: 15; Díaz-Montón, 2006: 3).

En resumen, la localización aparece en los 80 aplicándose solo al *software* y con la evolución de Internet en los 90 y los 2000, las compañías identifican la necesidad de adaptar los contenidos de sus productos y sitios web a los diferentes *locales*. Esto contribuyó a una expansión de la web y al desarrollo de programas de *software* más avanzados, los cuales se lanzan y distribuyen de manera simultánea al original en todos los mercados o *locales* de llegada.

### 2.2.2. Definición

La localización como tal es una disciplina muy reciente y, por este motivo, todavía no hay una definición universal de este concepto. Su denominación siempre se ha abordado desde diferentes perspectivas y, en este apartado, se proporcionarán definiciones propias a partir de las diferentes ideas de autores ya consagrados en materia de localización.

Para la gran mayoría de autores, la localización consiste en adaptar lingüística y culturalmente aquellos productos informáticos, multimedia o sitios web, así como la documentación que los acompaña al *locale* de llegada, para que estos se puedan comercializar de manera que dicho *locale* meta perciba el producto como propio u original de su área geográfica (Arevalillo-Doval, 2000: 01; Bermúdez-Bausela, 2004: 02; Corte, 2002: 01; Esselink, 2000: 11; Fernández-Costales, 2010: 25; Jiménez-Crespo, 2009: 03; Pérez-Fernández, 2010: 63; Sandrini, 2008: 02). Además,

esta adaptación responde a las necesidades de las personas que forman parte de un mismo *locale* meta (Esselink y Cadieux, 2002: 02).

La adaptación al *locale* de llegada se debe hacer correctamente y el producto final se debería llegar a percibir como original en el mercado de llegada. Si no se hace así, el texto no sería funcional en el *locale* meta (Jiménez-Crespo, 2008: 15). Según Jiménez-Crespo (2009: 10):

It should be noted that from a functionalist perspective, the substitution or adaptation of the conventions in the source text for those in the sociocultural context of reception is not automatic, but depends on the *skopos* of the translation and the norms of the target culture (Reiss y Vermeer 1984: 194). Nevertheless, it is logical to assert that most localised genres need to be functional texts in the target culture, and in principle all target texts should incorporate whichever conventions are established in the receiving locale.

De hecho, Jiménez-Crespo (2009: 10) afirma que, si los textos localizados son similares a otros textos escritos originalmente en el *locale* meta, se considerará que estos son de gran calidad.

Si se aborda la localización desde una perspectiva empresarial, se puede concluir que esta podría ayudar a las compañías a aumentar sus beneficios, ya que les posibilita vender sus productos en muchos más países (Anastasiou y Schäler, 2010: 09). Si se lleva a cabo correctamente, permitirá superar barreras legales, lingüísticas y culturales (Fernández-Costales, 2009: 03). Si, por el contrario, se obvia la localización en el proceso de internacionalización de un producto, podrían llegar a producirse pérdidas millonarias para los comerciantes del producto en cuestión (Cámara de la Fuente, 2005: 56).

Según O'Hagan y Ashworth (2002), la localización va más allá del transvase lingüístico, por lo que utilizan el concepto de «culturización» puesto que la cultura incluye algunas convenciones que no son solo lingüísticas. De esta manera, se da a entender que la localización culturiza el mensaje. La localización hace también referencia a la relación entre una lengua y el territorio o *locale* donde se habla. Esto implica que los productos deberán traducirse al idioma o idiomas que se hablen en dicho *locale* o territorio, no es suficiente con que el producto se comercialice en inglés (Diéguez-Morales; 2008: 02).

En esta culturización del mensaje, se distinguen tres niveles de adaptación (Jiménez-Crespo, 2008: 55; O'Hagan y Ashworth, 2002: 74):

- De primer nivel o habilitación (*enablement*): se crean o traducen documentos a las lenguas meta, pero la interfaz sigue estando en lengua origen (LO).

- De segundo nivel o localizado (*localized*): se traducen tanto la documentación como la interfaz, pero algunas de las herramientas del programa se dejan en LO.
- De tercer nivel o adaptado (*adapted*): se traduce y se adapta todo al mercado o *locale* de llegada.

Siguiendo esta teoría, O'Hagan y Ashworth (2002: 74) afirman que, en función del mercado o *locale* de destino a quien se dirija el producto, se encontrarán distintos niveles de adaptación. Cuanto más grande o importante sea el mercado o el *locale* meta, mayor será el nivel de adaptación necesario y, por tanto, también la inversión en localización (Pym, 2010: 11), ya que el producto localizado se orientará al destinatario (Bernal-Merino, 2013: 126; Fernández-Costales, 2010: 27). Esto es así porque la localización es también una estrategia de marketing y publicidad (Fernández-Costales, 2010: 30) no solo un proceso de adaptación lingüística.

Pym (2012: 214-215) sugiere dos formas distintas de orientar los proyectos de localización:

- Centralización de los contenidos multilingües: el producto tendrá la misma apariencia en todos los idiomas de llegada, lo que resultará en la estandarización extrema de los proyectos.
- Adaptación del contenido a las normas y gustos de cada *locale* o diversificación descentralizada.

El proceso de localización cuenta con varias fases y Esselink (2000: 414) elabora una lista de aquellas actividades que se suelen asociar con un proyecto de localización típico:

1. Preparación.
2. Traducción y revisión.
3. Ingeniería y testeo.
4. Localización de elementos gráficos.
5. Maquetación del documento o archivo.
6. Gestión de proyecto.

Scholand (2002: 05) identifica las diferentes tareas de la localización y sugiere dividir las en localización técnica, localización nacional y localización cultural. En primer lugar, la localización técnica hace referencia a la adaptación del sistema operativo, la conversión de doble byte, etc. La localización nacional representa la adaptación de símbolos y alfabetos. Por último, la localización cultural es la traducción del *software*, de la documentación y de los demás materiales que acompañan al producto.

Siguiendo con las definiciones de localización, algunos autores han defendido la denominación de «localización» ante términos como «adaptación», «naturalización» o «nacionalización», puesto que consideran que son muy poco específicos y no expresan bien el verdadero concepto (Arevalillo-Doval, 2000: 02; Pagans, 2002: 03). A estas, Esselink (2003: 21) añade que la localización combina la lengua con la tecnología para crear un producto final que sobrepase barreras tanto lingüísticas como culturales. En esta línea, Sandrini (2008: 01) agrega que la localización no se puede entender sin tecnología, ya que esta está muy ligada a los productos multimedia digitales y al trabajo con ordenador.

Por su parte, la desaparecida LISA (Localization Industry Standards Association) (1998: 03) propuso la definición «localization involves taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold». No obstante, LISA cambió su definición clásica a «localization is the process of modifying products or services to account for differences in distinct markets [...]» (Fry, 2007: 11). Aunque, no obstante, Pagans (2002: 03) añade que este término está ya muy implantado por lo que no merece la pena buscar uno nuevo, aunque se podría utilizar simplemente el hiperónimo «traducción» para evitar una mayor problemática.

La localización es un proceso unido intrínsecamente con el proceso globalización de un producto, por eso, O'Hagan y Ashworth (2002: 66-67) definen la localización como un proceso que facilita la globalización, mediante la superación de barreras lingüísticas y culturales específicas al receptor del producto, el cual no comparte el mismo entorno cultural y lingüístico que el creador o emisor del mensaje o producto. Para Jiménez-Crespo (2009: 01) la localización es un proceso tecnológico, textual, comunicativo y cognitivo mediante el cual textos digitales interactivos se adaptan para su uso en contextos culturales y lingüísticos diferentes.

Quah (2006: 19) define la localización como «the process of changing the documentation of a product, a product itself or the delivery of services so that they are appropriate and acceptable to the target society and culture». De esta manera, Jiménez-Crespo (2008: 14) considera la localización como parte de un proceso mucho más amplio que también incluye la globalización y la internacionalización y que, además es diferente en función del grado de localización.

Una vez establecidas las diferentes definiciones del término «localización», es momento de observar qué lo caracteriza. En el apartado siguiente, se resumirán las principales particularidades de esta disciplina, lo que permitirá discernir en qué aspectos difiere de la traducción e identificar qué puede aportar al tema que nos ocupa en la presente tesis.

### 2.2.3. Características de la localización

La localización cuenta con ciertas singularidades que la diferencian de otras disciplinas como la traducción. En el presente apartado, se expondrán algunas de sus principales características, según la bibliografía existente en materia de localización.

Para Scholand (2002: 01) la localización cuenta con las siguientes características:

- Corta vida de los productos.
- Diversidad de géneros y ámbitos.
- Distintas cualidades que el localizador debe tener.

En esta línea, Estellés (2010: 04) añade que la gran variedad de ámbitos y géneros y el lenguaje utilizado en cada uno de ellos será distinto en función de la situación comunicativa, lo que puede dar lugar a una gran variedad de géneros textuales como el técnico, el médico o el jurídico, entre otros muchos. Otra característica de la localización es que el objeto de trabajo en esta disciplina no son textos como tal, sino productos (Jiménez-Crespo, 2008: 39; Pym, 2012: 192). El énfasis está ahora en el producto y en el *locale* de destino, por eso, los localizadores pueden modificar el texto y no ser totalmente fieles al original para que el producto se adapte al mercado o *locale* de destino (Bernal-Merino, 2013: 125). Schäler (2008: 196) y Estellés (2010: 4) se suman a la afirmación de otros autores de que los materiales que se traducen en localización no solo están formados por textos, sino también por gráficos, audio, video, animaciones o fragmentos de código, entre otros. Además, según el tipo de producto y los materiales que lo formen, los localizadores deberán utilizar una determinada herramienta de traducción u otra (Schäler, 2008: 196).

Es más, la localización no solo se centra en los aspectos lingüísticos del texto traducible, sino que también tiene en cuenta la percepción de los distintos *locales* de los símbolos (monedas, unidades de medida, etc.), los iconos, los colores y demás efectos visuales (Anastasiou y Schäler, 2010: 02; O'Hagan y Ashworth, 2002: 10; Pagans, 2002: 01; Quah, 2006: 20; Ressin, 2015: 01). No tener en cuenta estos aspectos daría como resultado final una traducción no funcional para el *locale* de llegada (Jiménez-Crespo, 2008: 15).

Como se ha podido observar, la localización va un paso más allá que la traducción, ya que esta atiende aspectos más técnicos. Además, los lingüistas que se quieran dedicar a la localización requieren una formación especializada. Para observar, cómo entienden los estudiosos de la localización esta disciplina frente a la traducción, en el apartado siguiente, se ha realizado una comparativa entre traducción y localización.



#### 2.2.4. Traducción vs. Localización

La localización es una disciplina, cuya denominación es controvertida. Los diferentes autores que hablan de localización defienden teorías distintas sobre lo que es la localización. Algunos creen que la localización y la traducción son lo mismo, otros creen que son dos disciplinas totalmente distintas y separadas, pero otros creen que la localización es simplemente un paso dentro del proceso de traducción. Por este motivo, en el presente apartado, se comparan estos dos términos para entender exactamente en qué difieren entre sí ambas disciplinas. Se observará que algunos autores opinan que ambas forman parte de un mismo concepto, otros creen que una está comprendida dentro de la otra y otros que son conceptos totalmente distintos.

Para García-Saavedra Valle (s.f.: 413), la localización es una especialización dentro de la traducción. Del mismo modo, García-Nevarés (2016: 67-68) cree que los procesos de traducción y localización están entrelazados, por lo que no merece la pena hacer ninguna distinción entre ellos (García-Nevarés, 2016: 67-68). Hay autores como Diéguez-Morales (2008: 03) que directamente utilizan el término «traducción/localización».

Fernández-Costales (2009: 07) defiende que ambas disciplinas tienen el objetivo de transferir un texto de una cultura a otra. No obstante, cree que la localización implica también un aspecto mucho más técnico e informático que la traducción tradicional (Fernández-Costales, 2010: 52). Pese a estas diferencias la localización y la traducción tampoco deberían considerarse como disciplinas totalmente independientes porque las dos tienen la función común de adaptar un producto a un *locale* determinado (Fernández-Costales, 2010: 59). Otros autores, como Pérez-Cabrera (2012: 02), afirman que se pueden utilizar ambas denominaciones, es decir, utilizar tanto «localización» como «traducción», aunque la autora se decante finalmente por el término «localización».

Por otro lado, existen aquellos autores que opinan que la traducción es un hiperónimo de localización. Por un lado, Pagans (2002: 05) piensa que la localización es tan solo una modalidad de traducción, puesto que, en la traducción, siempre hay un grado de adaptación. De la misma forma, Pym (2008: 219) considera que la localización es solo una parte dentro del proceso de traducción. Por el contrario, la industria considera que la traducción es un proceso englobado dentro de la localización (Ressin, 2015: 21), visión que comparte Plaza-Lara (2014: 24), aunque esta defiende que ambos procesos son interdependientes.

Para otros autores como Bernal-Merino (2013: 125), la localización engloba algo más que la traducción de un producto, lo que implica muchas más tareas con un carácter más técnico. Algunas de estas tareas son la gestión de proyectos, el diseño y testeo del *software*, conversión de formatos y gestión de memorias de traducción y demás (Esselink, 2000: 02; Fernández-Costales, 2009: 04). La distinción de estas dos disciplinas también comprende el tipo de productos que se trabajan en cada una de ellas. Por ejemplo, en localización se trabaja con productos web, *software*, videojuegos o productos multimedia, aplicaciones de teléfonos móviles, tabletas, etc. (Plaza-Lara, 2014: 21-22).

Para Pagans (2002: 01), la localización es un proceso mucho más amplio que la traducción. La localización cuenta con más fases y requiere cierto grado de adaptación del contenido. Además, los profesionales de la localización necesitan tener unas habilidades técnicas que no son necesarias en el proceso de traducción. Aunque Cadieux y Esselink (2002) también defienden que traducción y localización son disciplinas distintas, creen que no llegarán a estar totalmente separadas, sino que los profesionales de la traducción deberán evolucionar y adquirir habilidades más técnicas que les permitan ejercer como localizadores.

Sandrini (2008) destaca además diferencias importantes en cuanto a las estrategias empleadas a la hora de traducir un texto. Por un lado, en traducción se puede optar por extranjerizar o domesticar un texto con una gran carga cultural. Por otro lado, al localizar un producto, este deberá adaptarse siempre al *locale* de llegada, por lo que se optará por una estrategia domesticante. No se puede seguir la misma técnica que en traducción, puesto que, si se extranjeriza demasiado el producto, los receptores de dicho producto podrían percibirlo como extranjero (Sandrini, 2008).

El tipo de archivos también difiere en localización. En esta disciplina, no solo se trabaja en formato texto, sino que se reciben archivos de diferentes extensiones como .exe, .dll, .rc o en lenguajes de programación para cuya lectura se suelen necesitar herramientas específicas (Sandrini, 2008). Destacamos también el hecho de que en localización el producto adaptado se suele comercializar de manera simultánea (*simship*), mientras que en traducción esta se hace cuando el producto original está ya terminado (Bermúdez-Bausela, 2004: 08; Esselink, 2000: 443; Fernández-Costales, 2010: 32). Además, en localización, suelen surgir muchas dificultades por la limitación de caracteres, la cual obliga a que los contenidos se sinteticen mucho más (García-Saavedra Valle, s.f.: 410).

A modo de conclusión, sería interesante citar la afirmación de Pym (2012: 224) sobre localización:

[...] la ideología de la localización manifiesta un interés activo en la defensa de la diversidad lingüística y cultural, así como en el fortalecimiento de mercados locales, ya que es ahí donde se venden y utilizan los productos. Cuanto más acceso tenga la localización a tecnologías de gran potencia, mayor será su influencia en el futuro de la diversidad. En lugar de extender un régimen uniforme, podrá ayudar a salvar las diferencias lingüísticas y culturales.

Según este autor, la localización juega un gran papel a la hora de preservar las identidades culturales, además ayuda a fortalecer los mercados de cada *locale* donde se comercialice el producto o productos en cuestión. Añade, además, que la localización favorece la diversidad cultural y lingüística de cada *locale* y que, por este motivo, los estudiosos de la traducción más teórica y tradicional deberían conceder importancia a la localización como disciplina de estudio (Pym, 2012: 224). Además, con la evolución de Internet y los contenidos digitales, la localización se ha convertido en una de las salidas profesionales con más posibilidades de futuro para los traductores. Sin embargo, García-Saavedra Valle (s.f.: 409) cree que la poca tradición de estudio académico y la escasez de fuentes de referencia representan un gran reto para esta disciplina.

Tras este breve recorrido por los orígenes, características y fundamentos de la localización, se han podido observar los diferentes puntos de vista que existen en torno a esta disciplina y las diferentes definiciones que se han proporcionado hasta la actualidad. Pese a que, en la actualidad, todavía no existe una definición unánime de lo que constituye la localización, todos los autores coinciden en que esta no debe pasarse por alto en los procesos de internacionalización y globalización, puesto que, si se realiza de manera correcta puede aportar grandes beneficios a las compañías.

Una vez expuestas las características y las definiciones de la localización, a continuación, se va a profundizar en la traducción automática. Se abordarán las principales definiciones aportadas hasta la fecha, además de los diferentes tipos de motores de traducción automática que han surgido a lo largo de los años y, por último, se realizará un recorrido histórico en que se dará cuenta de la evolución de estos sistemas.

### 2.3. Traducción Automática

Con el propósito de entender la traducción automática (TA) en todas sus variantes, se han estudiado sus orígenes y evolución, así como sus tipologías y definiciones, y estos se han expuesto en el presente apartado. Para comenzar esta sección, se ha hecho referencia a las diferentes definiciones ofrecidas en algunos de los estudios de investigación en TA que existen hasta la actualidad. Una vez descritas las principales características de la TA, se han enumerado y explicado los principales tipos de TA existentes hasta la fecha, desde sus inicios como TA basada en reglas, hasta la más novedosa traducción automática neuronal (TAN), los mecanismos de atención o el modelo *Transformer*. Dentro de esta, se hablará del tan necesario proceso de posesición de textos y, por último, se abordará la historia de la TA desde sus orígenes con Troyanskij y Artsrouni (Hutchins, 2001: 09) hasta la actualidad.

Aunque estos sistemas aparecen aproximadamente en los años 30, es en los años 50, a partir del *Imitation Game* de Allan Turing, con Weaver y la Rockefeller Foundation, entre otros, cuando la investigación en TA experimenta un auge (Hutchins, 1995: 435; Weaver, 1955). Aparecen por primera vez los sistemas basados en reglas, que posteriormente se fueron dejando de lado para dejar paso a los sistemas basados en corpus (TA estadística y TA basada en ejemplos), los sistemas híbridos y los sistemas de TA neuronal, entre otros. A continuación, en la Ilustración 5, se ha elaborado un esquema con los distintos tipos de TA contemplados en la presente tesis, a lo largo del tiempo:



Ilustración 5. Evolución cronológica de la TA (Elaboración propia).

Es necesario puntualizar que, aunque hayan aparecido nuevos sistemas de TA que son los que ocupan hoy en día la mayor parte de la investigación, las fechas son aproximadas y se basan en la bibliografía estudiada, por lo que los años indicados no indican la desaparición de dichos sistemas de TA o de las investigaciones sobre estos.

De forma resumida, se podría decir que la historia de los motores de TA como se conocen en la actualidad comienza alrededor de los años 60 con la TA basada en reglas (auge a finales de los años 80). Los motores de TA basados en reglas incorporan diccionarios bilingües para cada par

de lenguas, compuestos por términos en LO junto con sus correspondientes traducciones en LM. A estos se debe sumar la utilización de un *software* que aplica reglas gramaticales y lingüísticas para la creación de las oraciones en LM.

A principio de los años 90 cambia el paradigma de TA para centrarse en el desarrollo de sistemas de TA basados en corpus, los cuales analizan ejemplos de traducción en corpus paralelos creados por lingüistas, formados por gran cantidad de textos en que se han alineado las oraciones en LO con sus equivalentes en LM. Dentro de la TA basada en corpus, existen dos modalidades: la TA estadística (TAE), fundamentada en frecuencias y en modelos estadísticos de traducción extraídos de corpus paralelos, en que se busca la traducción más probable para determinada oración en LO; y la TA basada en ejemplos (EBMT o *example-based machine translation* en inglés), cuya base es la imitación o la explotación de traducciones similares (ejemplos) extraídas de corpus bilingües alineados. Aunque ambos modelos están incluidos dentro del paradigma de la TA basada en corpus, conviene distinguir la TA basada en ejemplos y la TAE, ya que funcionan de formas distintas. Se profundizará más sobre este tema en apartados posteriores de esta tesis.

Posteriormente, las investigaciones comenzaron a centrarse en la llamada TA híbrida, que combina características de los diferentes sistemas, principalmente TAE con TA basada en reglas, para aprovechar las ventajas de cada método y compensar las deficiencias de cada uno de estos modelos de TA. Sin embargo, desde 2016, año en que surgió la TA neuronal, comenzaron a aparecer estudios sobre sistemas híbridos que incluían el enfoque neuronal combinado con otros paradigmas de TA.

Con el tiempo, se comienzan a abandonar los enfoques híbridos y los sistemas de TAE para centrar las investigaciones en los modelos de TA neuronal, basados en redes neuronales y aprendizaje automático (*machine learning*). Estos sistemas cuentan con un tipo de aprendizaje de extremo a extremo (*end-to-end*), que permite traducir de forma fiable sin realizar muchos cambios al algoritmo del sistema principal, algo muy beneficioso para entrenar sistemas multilingües.

Los sistemas de TA neuronal también utilizan corpus paralelos (como en los sistemas basados en corpus), pero en estos sistemas se aplican modelos neuronales, en que las palabras se representan numéricamente a través de vectores y, para procesar las oraciones, se suelen utilizar redes neuronales recurrentes (RNN o *Recurrent Neural Networks* en inglés). Los sistemas de TAN, conformados por RNN, basados en un codificador (*encoder*) y un decodificador (*decoder*), podrían explicarse como una red neuronal que calcula la probabilidad de que una oración en LM sea la traducción de una oración en LO. Otros enfoques más novedosos, ya no

utilizan RNN, sino redes neuronales convolucionales (*CNN* o *Convolutional Neural Networks*). En este caso, en vez de generarse una codificación de la oración original completa a partir de representaciones de palabras sueltas en LO, el decodificador genera representaciones de cada palabra teniendo en cuenta el conjunto de palabras a su alrededor.

Como se ha podido comprobar a través de esta breve introducción a la evolución de la TA, dentro del mundo de la TA existe una gran cantidad de siglas y nomenclaturas, algunas de las cuales proceden de ramas más técnicas, ya que la TA se ha estudiado tanto desde el área de la lingüística, como desde la ingeniería y la informática. Por este motivo, y para una mejor comprensión de este apartado, se ha elaborado una tabla con las principales siglas que aparecerán en las siguientes secciones (Tabla 4).

Tabla 4. Tabla de siglas comunes en TA (Elaboración propia).

CNN	Red neuronal convolucional o <i>Convolutional Neural Network</i>
FAHQT	Traducción totalmente automática de alta calidad o <i>Fully Automatic High Quality Translation</i>
GNMT	Traducción automática neuronal de Google o <i>Google Neural Machine Translation</i>
Herramientas TAO	Herramientas de traducción asistida por ordenador
LM	Lengua meta
LO	Lengua origen
ML	Aprendizaje automático o <i>machine learning</i>
MT	Memoria de traducción
PE	Posedición
RNN	Red neuronal recurrente o <i>Recurrent Neural Network</i>
TA	Traducción automática
TAE	Traducción automática estadística
TAN	Traducción automática neuronal
TM	Texto meta
TO	Texto origen

La importancia del estudio de la TA radica, por un lado, en la necesidad de comunicación en una sociedad cada vez más globalizada en que la cantidad de información que circula por las redes es ingente y necesita ser traducida para que más gente pueda acceder a ella. Por otro lado, es imprescindible comprender el funcionamiento de estos sistemas de TA, para poder llevar a cabo una estrategia de posedición que aborde los problemas lingüísticos que la TA todavía no puede solucionar de manera totalmente autónoma, ya que, en la actualidad, los sistemas de traducción automática aún no han alcanzado la perfección. De esta forma, para entender el alcance de la TA, en el apartado siguiente, se han proporcionado las definiciones de TA observadas a lo largo de la historia en los estudios académicos de este ámbito.

### 2.3.1. Definición

Como se observará en futuros apartados de esta sección, la TA es una disciplina investigada a lo largo de los años no solo por lingüistas, sino también por informáticos y matemáticos, entre otros. De esta manera, es natural que hayan aparecido múltiples definiciones para este mismo concepto con distintos matices dependiendo del campo de estudio desde el que se investigue. A continuación, se ha hecho un compendio de las distintas definiciones que se han aportado desde los inicios de los estudios en TA, de manera que se podrá observar cómo se ha definido a lo largo de la historia.

Uno de los primeros autores en mencionar la TA es Warren Weaver en su memorándum de 1949, publicado en 1955 en el libro *Machine Translation of Languages: Fourteen Essays*. En dicho memorándum, Weaver (1955) llama a este tipo de traducción «traducción mecánica» o «traducción por ordenador». Para Weaver, la traducción entre lenguas, sobre todo en contextos internacionales como la UNESCO, representaba un problema, puesto que dificultaba la comunicación entre personas. Esta problemática es lo que le lleva a plantearse la posibilidad de que existiera una traducción mecánica llevada a cabo por ordenadores:

Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worthwhile (Weaver, 1955: 4).

De esta forma, según Weaver, la TA es un tipo de traducción que llevaría a cabo un ordenador especialmente diseñado para ello, el cual, pese a las dificultades léxicas, sintácticas y semánticas, podría traducir sin problemas materiales científicos que tuvieran un propósito meramente informativo, aunque el resultado fueran textos poco naturales, pero inteligibles.

En 1951, Bar-Hillel también denomina a este tipo de traducción «traducción mecánica», pero en 1960, este autor comienza a designarla con el nombre con que se conoce hoy en día, «traducción automática». Bar-Hillel (1951) distingue entre los siguientes tipos de TA:

- **TA total:** un tipo de traducción totalmente automática y sin intervención de un traductor humano profesional ni antes de introducir el texto en el mecanismo de traducción, ni después para revisar el texto resultante de dicha TA. Observa como principales problemas la falta de precisión y la ambigüedad semántica.

- **TA mixta:** se utilizaría en contextos donde las traducciones de los textos deban ser muy precisas. En este tipo de traducción participaría una máquina y uno o dos traductores humanos, los cuales deberían desempeñar su papel al inicio del proceso o tras finalizar la traducción.
- **TA con poseedor:** la traducción la lleva a cabo la máquina y el poseedor se encarga de editar el texto para eliminar las ambigüedades semánticas. Deberá ser capaz de producir un texto comprensible a partir del texto traducido mediante TA. Todo este proceso debería llevarse a cabo en menos tiempo del que un traductor humano sería capaz de traducir el mismo TO desde el principio.
- **TA con pre-editor:** el pre-editor debe organizar el texto que se va a traducir en un paso previo a la TA, de tal manera que se eliminen las ambigüedades gramaticales y semánticas para que el sistema de TA pueda producir un texto correcto y comprensible.

Bar-Hillel (1960: 94) hace referencia a la denominada «traducción totalmente automática de alta calidad» (Fully Automatic High Quality Translation o FAHQT) e indica que, en aquel momento, este tipo de traducción era muy difícil de obtener y que seguiría siendo así en el futuro.

Posteriormente, se siguió la misma línea y se definió la TA como un proceso de traducción llevado a cabo por un ordenador de un lenguaje natural a otro y sin la ayuda de un traductor humano, ni durante el proceso de traducción en sí, ni como poseedor del texto meta resultante que deberá ser comprensible y útil (Informe ALPAC, 1966: 19).

Hutchins (1995: 431) ofrece la siguiente definición de la TA:

The term 'machine translation' (MT) refers to computerized systems responsible for the production of translations with or without human assistance. It excludes computer-based translation tools which support translators by providing access to on-line dictionaries, remote terminology databanks, transmission and reception of texts, etc.

Hutchins (1995: 431) afirma además que los tipos de errores generados por los mecanismos de TA son diferentes de los errores que cometen los traductores humanos. Por último, excluye las herramientas TAO de esta definición, las cuales se utilizan como medios de asistencia a los traductores, no para realizar traducciones completas. Hutchins (2001: 08) completa esta definición y añade que la TA se puede comprender como «the efforts to develop computer programs (software) for the translation of natural languages, commonly and traditionally called 'machine translation' (MT), or, in non-English-speaking countries, 'automatic translation'».



Yunker (2006) comparte la definición de que la TA es el proceso mediante el cual se utilizan ordenadores para traducir texto sin intervención de un humano y detalla que los textos que se traducen mediante sistemas de TA, son aquellos que probablemente nunca hubieran sido traducidos por un traductor humano.

Además, Hutchins (2001: 08) precisa que, desde el principio, los investigadores se han concentrado en la TA de textos científico-técnicos, donde las dificultades o diferencias culturales son menores, donde la demanda de traducción ha sido mayor y donde no se necesitan traducciones perfectas estilística o gramaticalmente, sino traducciones que transmitan la esencia del texto original y que ahorren tiempo y dinero (Hutchins; 2001: 08; Hutchins, 2003: 09).

De esta manera, Hutchins (2003: 06-07) distingue entre cuatro principales tipos de TA:

- **TA de textos para diseminación:** los textos traducidos con sistemas de TA para este propósito deben ser revisados o poseditados por traductores humanos para alcanzar la calidad que este tipo de textos requieren. Otra posibilidad es que los TO se redacten en lenguaje controlado en un proceso previo a la traducción que facilite la tarea del sistema de TA.
- **TA para asimilación:** la calidad de los textos traducidos mediante sistemas de TA no es lo más importante. Es decir, puesto que los sistemas de TA todavía no producen traducciones perfectas, algunos usuarios prefieren obtener traducciones sin poseditar que les permitan entender más o menos el significado general del texto, que no tener ninguna traducción.
- **TA para intercambio:** propiciada por el vertiginoso paso al que aparecen contenidos en la web y se intercambian mensajes en línea. En estos contextos, la TA actúa de manera inmediata, aunque sus resultados puedan ser en ocasiones pobres. También en este contexto se están investigando sistemas de reconocimiento del habla que puedan traducir o interpretar en tiempo real el discurso oral.
- **Aplicaciones de TA integradas para acceso a la información:** por ejemplo, la integración de *software* de traducción en determinadas herramientas y sistemas de extracción de información en textos.

Por otro lado, Allen (2003: 301) distingue dos enfoques de TA según el nivel de posesición que requieran e identifica los siguientes:

**Inbound translation** (also referred to elsewhere as MT for acquisition or assimilation) is simply the process of translating to understand. **Outbound**

**translation** (also referred to as MT for dissemination), on the other hand, the process of translating to communicate.

Gaspari y Hutchins, (2007: 199) hacen referencia a los sistemas de TA en línea y los definen como los servicios a los que los usuarios pueden conectarse fácilmente en Internet, en los cuales se pueden introducir los textos o, directamente, las páginas web que se desee traducir.

Asimismo, Hutchins (2001: 05) considera que los sistemas de TA no son rivales para los traductores profesionales, pero que estos sirven como herramienta de asistencia en el proceso de traducción. Además, realiza una distinción entre la TA que lleva a cabo todo el proceso de traducción, aunque tenga que revisarse el texto resultante; las herramientas de traducción asistida por ordenador, que son una ayuda para el traductor profesional; y los sistemas de TA para el traductor no profesional, los cuales producen borradores de traducción que simplemente ayudan en la comprensión de los documentos. Hutchins (2001: 10) puntualiza lo siguiente:

What is still often forgotten (particularly by new researchers) is that MT is a practical task, a means to an end, and that translation itself (automated or not) has never been and cannot be 'perfect'; there are always other possible (often multiple) translations of the same text according to different circumstances and requirements. MT can be no different: there cannot be a 'perfect' automatic translation.

Es decir que no se puede calificar a la TA como imperfecta, puesto que la traducción (ya sea llevada a cabo por un humano o por un ordenador) de un texto siempre tendrá varias versiones posibles. Por este motivo, no se debería abandonar el uso de la TA para determinados contextos, puesto que se debería tener en cuenta los requisitos y el uso del texto concreto, para comprobar si, en última instancia, resulta rentable utilizar este tipo de sistemas frente a una traducción realizada por un traductor humano profesional.

Según Hutchins (2003: 10), para que la TA sea rentable, esta debe estar bien integrada en todos los procesos de documentación técnica de la compañía, desde la redacción de los textos técnicos hasta la publicación y distribución de dichos documentos. Además, el sistema de TA que se elija deberá ser capaz de traducir en diversos pares de idiomas para que sea verdaderamente útil y rentable.

Por último, tras estudiar el mercado de la TA en España, Torres-Hostench, Presas y Cid-Leal (2016: 01) proponen diferentes opciones para utilizar la TA:

- Motores comerciales de TA: estos se nutren con información propia de la empresa y, según la combinación lingüística para la que se usen, pueden alcanzar niveles de calidad satisfactorios. El inconveniente es que pueden resultar poco económicos y exigen una gran inversión de tiempo y recursos.
- Motores gratuitos: aunque presentan la gran ventaja de no tener coste, se nutren con información que la empresa misma deberá proporcionar, lo que presenta el inconveniente de la falta de confidencialidad.
- Alternativas mixtas: por ejemplo, el motor Moses, que es de acceso libre y código abierto, pero presenta el inconveniente de que también necesita una gran inversión de tiempo y recursos humanos para que haya realmente un aprovechamiento útil.

En su estudio, Torres-Hostench, Presas y Cid-Leal (2016: 01) concluyen que si los traductores reciben la formación necesaria podrán llevar a cabo las diferentes fases del proceso de TA como, por ejemplo, la preparación de la documentación, el entrenamiento del sistema de TA, la predicción y la posesión de los textos, así como la retroalimentación del sistema.

Kay (1997: 01), concluye que, aunque ha habido importantes avances en el mundo de la informática referentes al estudio de la TA, los avances en lingüística todavía no han llegado a solucionar los problemas que aún plantean estos sistemas, como pueden ser problemas semánticos, gramaticales y demás. Por este motivo, plantea la solución de que lo mejor es desarrollar sistemas de TA cooperativos, es decir, herramientas que ayuden a los traductores en su profesión.

A modo de conclusión, se podría afirmar que, en general, los autores están bastante de acuerdo en lo que representa la TA. No obstante, a lo largo del tiempo, han surgido modelos de TA diferentes con funcionamientos totalmente distintos unos de otros. Por este motivo, se considera que la definición de la TA está intrínsecamente relacionada con cada tipo de TA, sea cual sea. En parte, se podría pensar que esta diversidad de definiciones se origina por la misma evolución de estos sistemas, que han pasado de ser rudimentarias máquinas automatizadas y grandes ordenadores a sistemas personalizables y de aprendizaje automático (*machine learning*). De esta manera, todavía parece muy complicado ponerse de acuerdo en una definición única que englobe el significado de la TA en general. Así, en el apartado siguiente se describirán algunos de los principales sistemas de TA que han existido, con el propósito de adquirir una perspectiva más precisa de cómo funcionan y cómo han contribuido al desarrollo y la evolución de la TA.

### 2.3.2. Tipos de TA

Cuando se habla de traducción automática (TA), se suele pensar únicamente en modernos y complejos sistemas de traducción automática neuronal (TAN). No obstante, la TA se origina con una forma mucho más primitiva hace aproximadamente 90 años (Hutchins, 2004: 11) y, a partir de ahí, han ido surgiendo multitud de sistemas distintos que han evolucionado hasta la actualidad. De esta manera, para comprender exactamente qué es la TA y las tipologías existentes, a continuación, se han enumerado los principales tipos de TA que han documentados, sus características y algunas de sus principales ventajas e inconvenientes, empezando por la TA basada en reglas o *ruled-based machine translation* hasta llegar a la TAN, mencionando también los diferentes tipos de TA basada en corpus o *corpus-based machine translation*.

#### 2.3.2.1. TA basada en reglas

A finales de los 80, el paradigma de investigación sobre TA se basaba en reglas lingüísticas para análisis sintácticos, reglas para transferencia léxica o reglas morfológicas, entre otras. De ahí que el enfoque de la TA basada en reglas fuera el principal en aquel momento, con la aparición de sistemas como Ariane, Metal, Eurotra o Systran, entre otros. No obstante, a partir de los años 90, la investigación en TA dio un giro hacia sistemas basados en corpus, los cuales se han detallado en el apartado siguiente (Hutchins, 2015: 12).

Según Hutchins (2015: 20), la TA basada en reglas, comenzó a desarrollarse entre los años 1967 y 1989 en el contexto de las gramáticas o reglas de transferencia, así como el de las investigaciones y experimentos con sistemas de interlingua y basados en inteligencia artificial. Estos sistemas comenzaron siendo desarrollados con modelos que relacionaban directamente las cadenas de texto en LO con las cadenas de texto en LM, pero, poco a poco, empezaron también a relacionar representaciones sintácticas e incluso formas lógicas y representaciones semánticas (Cancedda et al., 2009: 26). Este tipo de sistemas basados en reglas utilizan datos probabilísticos y estocásticos<sup>7</sup>, además integran innumerables reglas lingüísticas e incorporan millones de diccionarios bilingües para cada par de lenguas (Hutchins, 1995: 435; Systran, s.f.).

---

<sup>7</sup> Una misma palabra u oración puede ser traducida de diferentes maneras por diferentes traductores e incluso por el mismo traductor (p. ej. puede variar el orden de las palabras). Si se entrena al sistema de TA de manera estocástica mediante un corpus con una misma oración ordenada de distintas formas, se le enseña, por ejemplo, que una misma oración puede traducirse de distintas maneras (Schulz, Aziz y Cohn, 2018).

Estos diccionarios bilingües se utilizan para entrenar al ordenador en cómo traducir determinadas oraciones, por lo que están compuestos por términos en LO con sus correspondientes traducciones en LM, para determinados pares de lenguas específicos, por ejemplo, español<>inglés. Dichos diccionarios se emparejan con *software* que aplica reglas gramaticales y lingüísticas como, por ejemplo, que un verbo vaya delante de determinado sustantivo. No obstante, según Yunker (2006: 09), incluso las reglas más exhaustivas pueden resultar insuficientes cuando, en las traducciones, aparecen palabras con usos diferentes. Además, añadir pares de lenguas nuevos o personalizar un par de lenguas ya existente supone mucho trabajo para los profesionales humanos que deben hacerlo, puesto que tienen que crear los recursos lingüísticos necesarios para que el sistema funcione correctamente (Somers, 2003; Yunker, 2006: 10).

Como se ha indicado anteriormente, los sistemas de TA basada en reglas funcionan mediante una serie de reglas de transferencia y generación morfológica, sintáctica y léxica, además se sirven de diccionarios y vocabularios bilingües o plurilingües (Sreelekha, 2017). Estos sistemas de reglas dependen enormemente del conocimiento lingüístico que contengan los diccionarios que conformen el sistema y de las reglas de desambiguación o de los corpus desambiguados manualmente, además son codificados por lingüistas profesionales, los cuales intentan describir el proceso de traducción (Somers, 2003).

Es decir, los sistemas de TA basada en reglas extraen información de extensos diccionarios y gramáticas bilingües o plurilingües, las cuales se complementan mediante reglas lingüísticas. Además, estos funcionan mediante un análisis o *parsing* del texto en LO y, basándose en la palabra u oración identificada en el TO, se crea una representación intermedia o simbólica (IR o *intermediate representation* en inglés), a partir de la cual se genera el texto en LM (Armentano-Oller et al., 2007: 05; Koletnik, 2011: 03; Sánchez-Martínez, 2008: 03).

El proceso que sigue la TA basada en reglas es complejo y, por este motivo, supone ciertas ventajas, pero también algunos inconvenientes, como se observa en la Tabla 4, a continuación:

Tabla 5. Ventajas e inconvenientes de la TA basada en reglas (Elaboración propia).

Ventajas	Inconvenientes
Gran control sobre las traducciones, ya que estas son resultado de reglas introducidas a mano (Lison, 2016).	Los resultados de las traducciones carecen de fluidez. (Systran, s.f.).

Los errores que cometen estos sistemas son más fáciles de predecir y poseer debido a que son muy repetitivos, lo que convierte a la TA basada en reglas en un buen sistema para la traducción de textos divulgativos (Sánchez-Martínez; 2008: 04).

La configuración personalizada de los diccionarios hace posible una mayor calidad de las traducciones y que el vocabulario utilizado se adecue a la terminología corporativa de cada empresa (Systran, s.f.)

Estos sistemas funcionan adecuadamente con *hardware* estándar, mientras que los sistemas de traducción automática estadística (TAE) requieren *hardware* más sofisticados (Systran, s.f.).

Se necesita invertir mucho tiempo y dinero para configurar el sistema de manera que alcance los estándares de calidad adecuados (Lison, 2016). Además, después de la configuración inicial del sistema, habrá que trabajar mucho no solo para mantenerlo, sino para que cada vez sea de mayor calidad (Systran, s.f.).

Incorporar nuevos pares de lenguas y nuevas reglas lingüísticas también supone una gran inversión de tiempo y esfuerzo (Sreelekha, 2017; Yunker, 2006: 10), ya que los lingüistas tienen que crear las entradas de los diccionarios y las reglas de transferencia manualmente (Bayatli et al., 2018: 49; Tyers y Donnelly, 2009: 64).

---

Dos de los principales sistemas de TA basada en reglas fueron Apertium y Systran. Por un lado, Apertium es una plataforma online de *software* libre y código abierto que traduce a diversos pares de lenguas entre idiomas como español, catalán, euskera, gallego, aragonés, inglés, francés, esperanto, occitano, portugués, rumano y bretón, entre otros. Apertium nació en 2004 del esfuerzo conjunto del Ministerio de Industria, Turismo y Comercio, de la Generalitat de Catalunya y de la Universidad de Alicante.

Al tratarse de una plataforma de código abierto permite crear un sistema basado en reglas, el cual, por su naturaleza, no requiere entrenamiento como tal. Además, hace posible que lingüistas voluntarios lo actualicen para mantener los pares de lenguas ya existentes e incorporar nuevos pares. Pese a que se creó para traducir entre pares de lenguas con raíces similares, como el español o el catalán, también se ha utilizado para traducir entre pares de

lenguas más distantes, como el español y el vasco (Armentano-Oller et al., 2006; Ortega, Lu, Meyers, y Cho, 2018: 213; Tyers y Donnelly, 2009: 57; Tyers, 2009: 214).

A continuación, la Ilustración 6 presenta la estructura del funcionamiento de Apertium según Forcada, Tyers y Ramírez-Sánchez (2009: 07):

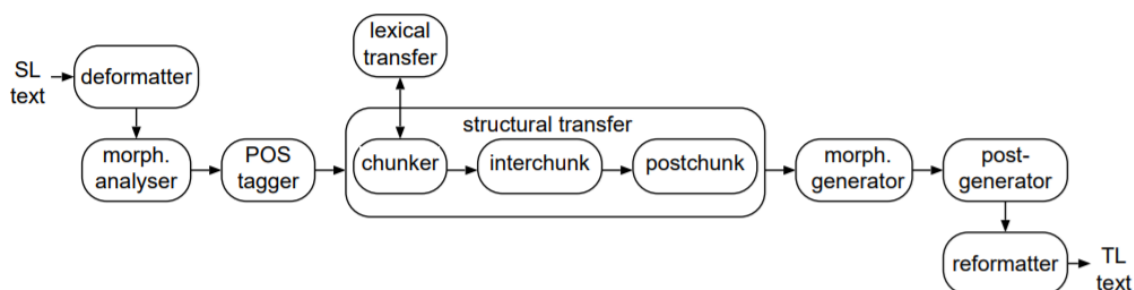


Ilustración 6. Funcionamiento de Apertium.

Como se observa, en la Ilustración 6, Apertium está formado por módulos encadenados que procesan el texto por etapas entre los cuales encontramos los siguientes (Armentano-Oller et al., 2007; Forcada, 2009: 19; Forcada, Tyers y Ramírez-Sánchez, 2009: 07):

- Desformateador: despoja al TO de su formato inicial.
- Analizador morfológico: segmenta el texto en LO en formas superficiales (FS), asigna a cada una de estas FS una o más formas léxicas (FL). Cada una de estas con un lema, una categoría léxica, e información de flexión morfológica.
- Desambiguador léxico categorial: POS tagger o *part-of-speech tagger* en inglés. Elige la FL correspondiente y correcta para cada FS ambigua.
- Módulo de transferencia léxica: los diccionarios proponen unidades léxicas para cada caso y par de lenguas.
- Módulo de transferencia estructural: propone estructuras sintácticas para cada patrón estructural (sintagma) en LO.
- Generador morfológico: genera a partir de la forma léxica en LM una FS correctamente flexionada.
- Postgenerador: realiza transformaciones ortográficas en LO como, por ejemplo, contracciones o apostrofaciones.
- Reformateador: da al TM el formato original que tenía el TO.

Por otro lado, Systran, del que ya se habló en apartados anteriores, nace en los años 70, de la mano de su creador, Peter Toma, como un sistema de TA basada en reglas muy potente y se consolidó como una herramienta comercial que sigue existiendo hoy en día, aunque, tras convertirse más adelante en un sistema híbrido que incorporaba funciones de la TAE (Gomes de

Oliveira y Anastasiou, 2011: 122), actualmente ya utiliza TAN y traduce a más de 140 pares de idiomas (Systran, s.f.). Systran incorpora recursos lingüísticos, entradas de diccionario (entre 100 mil y 800 mil entradas léxicas) y reglas de desambiguación para cada idioma o par de lenguas (Dugast, Senellart y Koehn, 2007: 221).

Además, según Dugast, Senellart y Koehn (2007: 221), el sistema de Systran se puede utilizar de dos formas distintas:

1. Motores de traducción específicos para empresas que necesiten traducir y poseer documentos de campos mayoritariamente técnicos. Con posibilidad de incorporar diccionarios y memorias de traducción propios.
2. Como aplicación para el escritorio a través de *plugins* o en herramientas de posesión.

Uno de los mayores logros de Systran llegó en 1997, cuando la compañía firmó un acuerdo con el motor de búsqueda AltaVista, a partir del cual aparecería Babel Fish, un servicio de traducción automática gratuito al cual podía acceder cualquier usuario de Internet (Gaspari y Hutchins, 2007: 200).

Pese a que la TA basada en reglas ofrece ciertas ventajas, como su óptimo funcionamiento en *hardware* estándar, que no necesite entrenamiento o unos resultados relativamente aceptables, uno de los principales inconvenientes que presenta es que, para algunos pares de lenguas menos conocidos, puede haber falta de recursos y métodos apropiados, lo que no permitiría la desambiguación ni la elección de léxico apropiado (Tambouratzis, Vassiliou y Sofianopoulos, 2016: 132).

De esta manera, el paradigma en aquel momento, que se había centrado en el desarrollo de este tipo de sistemas basados en reglas y diccionarios, cambió a principios de los años 90. A partir de entonces, las investigaciones en TA empezaron a centrarse en los sistemas de TA basados en corpus de los cuales formaban parte la TA estadística (TAE) y la TA basada en ejemplos (EBMT o *example-based machine translation* en inglés) (Hutchins, 2007: 40; Tripathi y Sarkhel, 2010: 390), las cuales se describirán en el apartado siguiente.

#### 2.3.2.2. *TA basada en corpus*

Como se ha observado en el apartado anterior el paradigma de investigación en TA hasta finales de los 80 se basaba en estudiar la TA basada en reglas. No obstante, a partir de 1989, el predominio de la investigación en TA basada en reglas desaparece, una situación que fue propiciada por la aparición de nuevos métodos y estrategias basados en corpus (Forcada, 2009: 14; Hutchins, 2005: 63; Koletnik, 2011: 03).



En este nuevo enfoque, denominado TA basada en corpus o *corpus-based machine translation*, el sistema de TA analiza ejemplos de traducción en corpus paralelos elaborados por lingüistas expertos. Estos corpus paralelos están compuestos por un gran número de textos bilingües en que se han alineado las oraciones en LO con sus equivalentes en LM, a partir de los cuales el mecanismo de traducción aprenderá a traducir. De esta manera, la calidad de la TA basada en corpus depende de la disponibilidad de textos paralelos, alineados a nivel oracional, para determinados pares de lenguas (Armentano-Oller et al., 2007: 05; Dandapat y Lewis, 2018: 109; Forcada, 2009: 14-15; Poncelas et al. 2018: 249; Sánchez-Martínez, 2008: 03). Costa-Jussa et al. (2012: 247) afirman lo siguiente:

[...] under the corpus-based approach the knowledge is automatically extracted by analysing translation examples from a parallel corpus built by human experts. The advantage is that, once the required techniques have been developed for a given language pair, MT systems should – theoretically – be quickly developed for new language pairs using provided training data.

Por lo tanto, se puede afirmar que los sistemas de TA basada en corpus utilizan ejemplos de traducción extraídos del corpus empleado para entrenar el mecanismo y que han sido recogidos en los corpus paralelos, los cuales están alineados en determinados pares de lenguas. Estos ejemplos son los que el sistema de TA utilizará para deducir la traducción más probable correspondiente a cada oración del TO.

La TA basada en corpus presenta ciertas ventajas con respecto al modelo basado en reglas. Por ejemplo, que una vez se han desarrollado las técnicas necesarias para un par de lenguas determinado, en teoría resulta más fácil adaptar los sistemas a otras lenguas ya que no hay que diseñar reglas específicas para cada nuevo par de idioma que se quiera insertar en el sistema, como pasaba en los sistemas de TA basada en reglas. En los sistemas de TA basada en corpus, las reglas lingüísticas, las cuales proporcionan información semántica y sintáctica, están implícitas en los textos que conforman el corpus (Costa-Jussa et al., 2012: 247; Tambouratzis, Vassiliou y Sofianopoulos, 2016: 132).

Precisamente, una de las principales críticas que recibe la TA basada en reglas, es el gran esfuerzo que supone actualizar el sistema cada vez que se quieren incluir nuevos pares de lenguas, ya que los lingüistas deben incluir nuevas reglas manualmente. Por el contrario, en los sistemas basados en corpus no se requiere tal esfuerzo, ya que estas reglas ya existen de forma implícita en el corpus. Sin embargo, según Bayatli et al. (2018:49), si no existe un corpus para determinado par de lenguas, solo habría dos opciones posibles «the first is to create a new

parallel corpus, either by translating millions of words from scratch (requiring effort from translators), or by finding parallel text online and processing it (requiring effort from programmers)» (Bayatli et al., 2018: 49).

Además, los sistemas basados en corpus requieren grandes cantidades de textos, oraciones y palabras alineados para que los corpus paralelos den resultados de traducción de alta calidad. Los pares de lenguas mayoritarios (como español-inglés o inglés-francés) sí que disponen de cantidades significativas de textos para corpus. No obstante, los pares de lenguas minoritarias no disponen de tantos recursos (ejemplos de estos pares de lenguas serían (francés-catalán o inglés-afrikáans, entre otros). Así, crear sistemas basados en corpus para estos pares de lenguas con menos recursos puede convertirse en una tarea muy ardua y en ocasiones casi imposible (Sánchez-Martínez, 2008: 05).

Los dos principales enfoques en que se divide la TA basada en corpus son la TA estadística (TAE) y la TA basada en ejemplos o *example-based machine translation*. La TAE se basa principalmente en frecuencias y combinaciones de palabras y, a grandes rasgos, la TA basada en ejemplos se basa en la extracción y la combinación de oraciones o textos cortos (Armentano-Oller et al., 2007: 05; Costa-Jussa et al., 2012: 247; Hutchins, 2005: 63; Koletnik, 2011: 03; Sánchez-Martínez, 2008: 03; Sreelekha, 2017; Tambouratzis, Vassiliou y Sofianopoulos, 2016: 132; Tertoolen, 2012: 143). Se hablará de cada una de ellas detalladamente a lo largo de esta sección.

En primer lugar, dentro de la TA basada en corpus, se encuentran los sistemas de TA estadística o TAE, los cuales utilizan modelos estadísticos de traducción extraídos de corpus paralelos monolingües o plurilingües. En el presente apartado se definirá la TAE y se analizarán las principales ventajas e inconvenientes que presenta este modelo de TA. En esencia, el método utilizado en este tipo de sistemas de TA es alinear oraciones, palabras sueltas o grupos de palabras en textos paralelos y estimar la probabilidad de que cierta palabra u oración en LO se corresponda con otra palabra u oración en LM, la cual se extrae de una de las traducciones del corpus en LM con que esté alineada (Armentano-Oller et al., 2007: 04; Seoane-Vicente, 2015: 219; Tertoolen, 2012: 144). A partir de estos corpus en LO y LM alineados, el sistema busca de forma probabilística la mejor traducción posible para cada elemento de la oración «partiendo del segmento más grande posible (frase completa) hasta el más pequeño (palabra)» (Sánchez-Gijón y Oliver, 2018). Según Brown et al. (1988: 71), el método de estos sistemas de TAE se basa en la disponibilidad de textos originales con sus respectivas traducciones, para formar corpus paralelos en distintos pares de lenguas.

Desde una perspectiva más técnica, se podría afirmar que la TAE es un tipo de tecnología basada en datos, es decir, que no solo cuenta con reglas y diccionarios como la TA basada en reglas, sino que también emplea gigantescos corpus paralelos en LO y LM. López (2008: 02) asegura que la TAE trata la traducción como un problema de *machine learning*, lo que significa que aplica un algoritmo de aprendizaje a una gran cantidad de texto ya traducido (conocido como corpus paralelo, texto paralelo bitexto o multitexto). El objetivo de la TAE será aprender a traducir a partir de estos corpus paralelos con fragmentos de texto alineado en LO y LM, a partir de un proceso estadístico basado en el análisis de datos y en la identificación de patrones (Cancedda et al., 2009: 02; Yunker, 2006: 11).

Los modelos estadísticos son modelos probabilísticos estimados a partir de un corpus paralelo, en que los sistemas buscan la traducción más probable para determinada oración en LO (Lison, 2016; Sánchez-Martínez, 2008: 03; Systran, s.f.). Según Koletnik (2011: 03), los sistemas de TAE dependen de una gran cantidad de traducciones realizadas por profesionales humanos que el sistema utilizará para inferir un modelo estadístico de traducción, con la idea de que para cada elemento del TO hay un número determinado de traducciones posibles. El sistema asignará la probabilidad más alta a la traducción que se considere más adecuada. Para comprender mejor el funcionamiento de los sistemas de TAE, se propone como ejemplo el funcionamiento del Traductor de Google cuando este todavía utilizaba TAE, tal como lo explican Gomes de Oliveira y Anastasiou (2011: 122):

Google's SMT system needs to be fed with large amounts of text to work efficiently and these texts must correspond to the language pair chosen by the user, or the system will start using the interlingua approach, which first translates source language to interlingua (mostly English) and then into a target language. (Gomes de Oliveira y Anastasiou, 2011: 122).

Para Sánchez-Gijón y Oliver (2018), la calidad de la traducción depende directa e indirectamente de la disponibilidad de un corpus de trabajo, por lo que pueden plantear una serie de ventajas e inconvenientes que se deben tener en cuenta a la hora de decidirse a usar este tipo de sistemas estadísticos.

Una de las principales ventajas de estos sistemas es que las reglas lingüísticas están implícitas en el corpus de textos, por lo que no habrá que escribir estas reglas a mano como requieren los sistemas basados en reglas, lo que supone menos trabajo manual y un ahorro de tiempo y dinero. Además, esto también evita tener que crear reglas nuevas cada vez que se quiera introducir un par de lenguas en el sistema (Seneff, Wang y Lee, 2006: 217; Tyers y Donnelly,

2009: 64; Yunker, 2006: 11). De la misma forma, si hubiera errores en el corpus, estos se podrían solucionar rápidamente realizando cambios en el algoritmo e incorporando nuevos textos al corpus para seguir mejorando la calidad del sistema (Yunker, 2006: 11).

Aunque los sistemas estadísticos produzcan traducciones que en ocasiones son correctas e incluso idiomáticas, una de sus desventajas es que suelen producir traducciones poco consistentes y con errores lingüísticos como, por ejemplo, errores en el orden en que se disponen los elementos que conforman las oraciones (Quirk y Menezes, 2005: 99; Systran, s.f.). Asimismo, para que se puedan aprovechar todas las ventajas que ofrece la TAE con respecto a otros tipos de TA, habrá que asegurarse de que el corpus sea de gran calidad, así como de una extensión considerable, a la hora de entrenar el sistema (Lison, 2016; Ramírez-Polo, 2012: 226; Seneff, Wang y Lee, 2006: 217). Aunque se pueda crear un sistema de TAE de forma más rápida, estos sistemas dependen de la existencia de corpus multilingües muy extensos, los cuales pueden incluso alcanzar los dos millones de palabras tanto para dominios específicos como para lenguaje en general (Systran, s.f.; Tertoolen, 2012: 144).

Las enormes cantidades de corpus que se necesitan para configurar estos sistemas pueden llegar a representar un problema, ya que muchos pares de lenguas minoritarios carecen de corpus con una extensión significativa para que los resultados que produzca la TAE sean de calidad. De igual manera, muchas empresas no disponen de estas cantidades de corpus multilingües tan considerables para crear los modelos de traducción. Por este motivo, las traducciones entre los pares de lenguas con más recursos serán mejores que aquellas entre pares de lenguas con pocos recursos. (Armentano-Oller et al., 2007: 04; Gomes de Oliveira y Anastasiou, 2011: 122; López, 2008: 40; Systran, s.f.; Yunker, 2006: 12).

Por último, aunque el entrenamiento del sistema mediante corpus de calidad sea automático y mucho más económico, si el corpus con que se entrena contiene textos en lenguaje general, cuando se quiera traducir textos especializados los resultados serán de poca calidad. Además, crear y gestionar un sistema de calidad requiere un *hardware* potente, ya que la TAE, consume mucha CPU para obtener un rendimiento promedio (Systran, s.f.).

La TAE cuenta también con algunas ventajas con respecto a la TA basada en reglas, ya que para desarrollar un sistema basado en reglas se puede tardar años, dado el gran trabajo que supone escribir las reglas y los diccionarios, pero en la TAE las reglas están implícitas en el corpus (Tyers y Donnelly, 2009: 64; Uszkoreit et al., 2009). Según Systran (s.f.), entre la TAE y la TA basada en reglas se encuentran también las diferencias indicadas en la Tabla 5, a continuación:

Tabla 6. Principales diferencias entre la TAE y la TA basada en reglas (Elaboración propia).

TAE	TA basada en reglas
Buena calidad si se dispone de corpus grandes.	Traducciones de calidad en ámbitos alejados del dominio o ámbito concreto.
La traducción no es previsible.	La traducción es previsible por naturaleza.
La traducción es fluida, por lo que se lee bien y, suele cumplir con las expectativas de los usuarios.	Poder personalizar los diccionarios garantiza una mejor calidad y el uso de terminología corporativa, pero el resultado pueden ser traducciones poco fluidas.
Entrenar los sistemas mediante corpus de calidad es automático y más económico. Pero el entrenamiento mediante corpus de lenguaje general y no de lenguaje específico suele ser deficiente.	El ciclo para personalizar el sistema para que este sea de calidad puede ser largo y costoso.
Se necesita una gran inversión en <i>hardware</i> para administrar modelos de traducción de gran tamaño.	Se consigue un alto rendimiento también con <i>hardware</i> estándar.

A modo de resumen, se podría decir que los sistemas de TAE basan sus traducciones en modelos estadísticos y que requieren una gran cantidad de textos en corpus paralelos. Estos textos se utilizan para estimar los parámetros de traducción que deberán seguir estos sistemas, los cuales sirven para deducir la traducción más probable para un determinado TO.

A continuación, se observará el otro modelo de TA dentro del paradigma de la TA basada en corpus, la traducción automática basada en ejemplos. Esta difiere de los modelos estadísticos en que se basa principalmente en la extracción y la combinación de oraciones. En el apartado siguiente, se entrará más en detalle sobre este aspecto y sobre las demás características, ventajas y limitaciones de este tipo de sistemas. Ong, Go, Morga, Nunez y Veto (2007: 17) describen el funcionamiento de este tipo de sistemas de la siguiente manera:

EBMT systems learn how a certain sentence is to be translated by being trained on a given bilingual corpus, which contains a set of sentences in the source language with a corresponding translation in the target language. Correspondences between the sentences are learned by the system and subsequently used in translation. Translations are generated by comparing the input from a source language to the examples or templates stored in a database, finding the best match possible before deducing the equivalent text in the target language.

Es decir, que los sistemas de traducción automática basada en ejemplos (EBMT o *example-based machine translation* en inglés) funcionan mediante la imitación o la explotación de traducciones similares que ya han sido traducidas anteriormente, para que el sistema traduzca de LO a LM las oraciones del TO. Para ello, se utiliza una gran cantidad de textos bilingües alineados almacenados en un corpus (Costa-Jussa et al., 2012: 247; Oliver, 2014: 19; Ong, Go, Morga, Nunez y Veto, 2007: 17; Tambouratzis, Vassiliou y Sofianopoulos, 2016: 132; Tomás-Gironés, 2003: 13).

Tomás-Gironés (2003: 13) describe el funcionamiento de la TA basada en ejemplos de la siguiente manera:

[...] se introduce un texto de entrada del que se pretende realizar una traducción. El sistema fragmenta esta entrada en “unidades de texto” y busca en la base de datos los ejemplos almacenados que coinciden en mayor o menor medida con estas unidades. En el caso de un sistema de memorias de traducción, toda la información localizada que se considere de interés será mostrada para que un traductor humano tome las decisiones oportunas y construya las frases de salida más adecuadas. La traducción basada en ejemplos trata de ir un poco más lejos, sustituyendo la participación del traductor humano por un mecanismo automático, que explota los ejemplos encontrados combinándolos y aventurándose a producir una sentencia de salida.

Este no es el único que relaciona la TA basada en ejemplos con las memorias de traducción (MT). Oliver (2014: 19) indica que «en cierto modo se puede ver como un caso extremo de uso de memorias de traducción», no obstante, lo que diferencia la TA basada en ejemplos de las memorias de traducción es que las decisiones de traducción en la TA basada en ejemplos las realiza un ordenador y no un humano como con las MT.

Sato y Nagao (1990: 01) afirman que, fundamentalmente, la TA basada en ejemplos «translates a source sentence by imitating the translation of a similar sentence already in the database». Es decir, que la TA basada en ejemplos es una traducción por analogía, en que, dada una oración

en LO, el sistema encontrará dentro de un corpus paralelo la mejor opción de traducción en LM para la oración en LO. La oración en LO se analiza por partes o sintagmas, cuyas traducciones se extraerán del corpus paralelo para formar las oraciones correspondientes en LM, a partir de *templates* o plantillas de traducción (Costa-Jussa et al., 2012: 247; Rubio y Bernal, 2016: 62; Sánchez-Martínez, 2008: 03; Tambouratzis, Vassiliou y Sofianopoulos, 2016: 137).

Estas *templates* o plantillas de traducción se definen en este contexto como pares de oraciones bilingües en que las unidades de traducción correspondientes (oraciones o palabras) se alinean y se sustituyen por variables. Estas plantillas contienen patrones alineados en LO y LM consistentes de variables y constantes. Los patrones en LO se utilizan como referencia para la oración que se va a traducir. Los patrones en LM se usan para generar la traducción final. Los patrones conservan el orden de las palabras de la traducción final, independientemente de la variación de las estructuras de las oraciones en LO y el LM (Ong, Go, Morga, Nunez y Veto, 2007: 18).

A modo de aclaración, se podría afirmar que la TA basada en ejemplos utiliza corpus bilingües paralelos (también conocidos como bitextos), los cuales normalmente están alineados a nivel oracional. El algoritmo se encarga de encontrar correspondencias entre las cadenas del texto que se quiere traducir en LO y las cadenas de texto en LO almacenadas en el corpus. A continuación, se buscarán en el corpus los segmentos en LM correspondientes, a partir de los cuales se realizará una traducción final en LM de la oración original en LO.

La TA basada en ejemplos enfatiza el aprendizaje mediante ejemplos (Och, 2005: 07), además se podría considerar este tipo de sistemas como una implementación del enfoque del razonamiento basado en casos del campo del *machine learning*, el cual se podría describir como un proceso en que se utilizan soluciones a problemas anteriores para resolver problemas nuevos (Costa-Jussa et al., 2012: 247)

Aunque los procesos principales de los sistemas basados en ejemplos (búsqueda de correspondencias, extracción y adaptación) la distinguen de otros paradigmas de TA, esta cuenta con muchos procesos secundarios que nos recuerdan a estos. Por ejemplo, el uso de diccionarios y de tesauros recuerdan a la TA basada en reglas. En el caso de la TA basada en ejemplos se utilizan para distinguir palabras polisémicas gracias al contexto o para reconocer variables como nombres, fechas o numerales a partir de coincidencias estructurales en cadenas procesadas de textos (Carl y Way, 2003: xix; Hutchins, 2005).

De esta manera también convendría distinguir la TA basada en ejemplos de la TAE, ya que ambas forman parte del paradigma de la TA basada en corpus, pero funcionan de formas distintas. Por

un lado, Hutchins (2005: 197) propone las siguientes diferencias entre la TA basada en ejemplos y la TAE:

[...] on the one hand, statistical machine translation (SMT), based primarily on word frequency and word combinations, and on the other hand, example-based machine translation (EBMT), based on the extraction and combination of phrases (or other short parts of texts).

Por lo tanto, la TAE se basa en la frecuencia y la combinación de palabras, pero la TA basada en ejemplos se basa en la extracción y la combinación de oraciones o secuencias de palabras (la unidad básica en TA basada en ejemplos son las oraciones o los fragmentos, no las palabras individuales).

Aunque en ambos casos se utilizan corpus paralelos bilingües de textos originales con sus correspondientes traducciones, en los sistemas basados en ejemplos, las nuevas traducciones se forman a partir de traducciones recopiladas anteriormente. La TAE, además de esto, también incluye un modelo probabilístico (España-Bonet y Costa-Jussa, 2016: 04). Este modelo probabilístico o estadístico indica la probabilidad de que un fragmento del TO en concreto concuerde con el fragmento en LM correspondiente (Oliver, 2014: 19; Rubio y Bernal, 2016: 62), por lo tanto, como indica Oliver (2014: 19) «el sistema es capaz de traducir a partir de una serie de parámetros que han sido calculados a partir de una gran cantidad de textos originales y sus traducciones».

Del mismo modo, Hutchins (2005: 201) indica que, lo que diferencia la TA basada en ejemplos de la TAE (y también de la TA basada en reglas) es que los procesos básicos de la TA basada en ejemplos se basan en analogías, o lo que es lo mismo, en encontrar oraciones o fragmentos en el corpus que sean similares al texto en LO de entrada que se quiera traducir, y en la adaptación y combinación de estas oraciones o fragmentos para producir oraciones o fragmentos en LM. Es decir, se basa en encontrar ejemplos análogos para encontrar las traducciones correspondientes del texto original en LO.

La TAE, por otro lado, no se basa en analogías, sino en correlaciones. Esta se fundamenta en un modelo de traducción estadístico aplicado a un corpus bilingüe paralelo, en el que se sustituyen palabras o fragmentos en LO por sus correspondientes traducciones en LM. Estos corpus han sido previamente alineados para establecer correlaciones entre los segmentos en LO y los segmentos en LM. Después de establecer estas correlaciones, se ordenan los fragmentos en LM siguiendo un modelo lingüístico monolingüe también preparado previamente (Hutchins, 2005: 201).



En resumen, se podría afirmar que la TA basada en ejemplos utiliza una serie de ejemplos de traducción recogidos en un corpus bilingüe como principal fuente de conocimiento. El proceso de traducción en este caso se basa en alinear los fragmentos en LO del texto que se va a traducir con los fragmentos en LO del corpus y, posteriormente, encontrar los fragmentos equivalentes en LM en el corpus como posibles traducciones. La TAE basa sus traducciones en modelos estadísticos que requieren grandes corpus paralelos con información relevante para que el modelo estadístico pueda determinar la mejor traducción para cada segmento en LO.

La TA basada en ejemplos presenta algunas ventajas con respecto a otros modelos de TA. Por ejemplo, estos sistemas se pueden desarrollar con menores cantidades de datos para su entrenamiento, pero, aunque el tamaño de la base de datos o corpus de ejemplos dependerá del sistema en cuestión y de los objetivos de los desarrolladores, un corpus grande reduce los problemas que pueda ocasionar la escasez de datos (Gough y Way, 2004). Además, la calidad de los resultados que ofrecen estos sistemas depende de la calidad de las traducciones de referencia, por eso, Way y Gough (2005: 03) afirman que se puede controlar los resultados de traducción si se controla la calidad de los textos de referencia utilizados para entrenar el sistema, algo que han relacionado también al uso de lenguajes controlados.

Pese a todas estas ventajas, la TA basada en ejemplos también plantea una serie de desventajas y dificultades. Anteriormente, se ha mencionado que la calidad de las traducciones de los sistemas basados en ejemplos depende de la calidad de los datos de entrenamiento, es decir, de los corpus. Aunque esto pueda representar una ventaja, también puede convertirse en una dificultad si no se dispone de datos de calidad. Por otro lado, Tomás-Gironés (2003: 14) plantea también la dificultad de determinación de la unidad de texto para alinear en los corpus paralelos que los forman. Es decir, determinar si la unidad que se va a alinear es una palabra, una oración o una suboración. Si se elige la oración, tal vez la unidad sea demasiado larga, pero, si se elige la suboración, habría que delimitar correctamente las unidades, si no las traducciones podrían ser incorrectas.

A modo de conclusión, se podría asegurar que la TA basada en ejemplos fue uno de los paradigmas dentro de la TA basada en corpus que más se utilizó en los inicios, no obstante, con la aparición y mejora de la TAE, esta se convirtió en el centro de estudio y desarrollo en TA, lo que provocó un descenso en el número de estudios centrados en investigar y desarrollar nuevos sistemas basados en ejemplos. Sin embargo, posteriormente, con la aparición y la evolución de la TA neuronal (TAN), se produjo un giro hacia la investigación de este nuevo paradigma de TA, el cual se ha explicado en detalle en el siguiente punto.

### 2.3.2.3. TA Híbrida

Tras la aparición de los métodos basados en reglas y aquellos basados en corpus (TA estadística y TA basada en ejemplos), surgieron nuevas metodologías y modelos de traducción que combinaban características de los diferentes sistemas para aprovechar las ventajas de cada método. Un ejemplo serían los sistemas que combinaban TAE con TA basada en reglas, el cual fue uno de los métodos híbridos más explotados. Hutchins (2005: 4) afirma que, aunque las metodologías basadas en sistemas estadísticos eran las predominantes, seguía habiendo algunos aspectos de la TA para los cuales se seguían utilizando metodologías de la TA basada en reglas, como, por ejemplo, la realización de análisis sintácticos para mejorar el proceso de reordenación de las oraciones entre lenguas de tipologías diferentes (p.ej.: inglés y japonés), el trato de lenguas ricas morfológicamente (p.ej.: ruso) o el problema de los nombres transliterados (p.ej.: chino), entre otros.

Por este motivo, las investigaciones empezaron a adoptar enfoques híbridos entre sistemas de TAE y de TA basada en reglas, ya que este tipo de sistemas híbridos ayudaban a compensar las deficiencias de cada tipo de enfoque de TA (España-Bonet y Costa-Jussa, 2016: 10; Sánchez-Martínez, 2008: 04; Uszkoreit et al., 2009). Ramírez-Polo (2012: 227) también advierte de la aparición de estos sistemas híbridos «recently, the emergence of new hybrid systems combining both rule-based methodologies with statistical algorithms has burst into the scene of MT». Para España-Bonet y Costa-Jussa (2016: 03) la TA híbrida implica «an integration of the system architectures».

Los sistemas estadísticos funcionan bien si se tiene un corpus paralelo considerable y de buena calidad, pero pueden mejorar sintácticamente y gramaticalmente si se integran características de los sistemas basados en reglas y si se incorporan normas lingüísticas durante las diferentes fases del proceso de traducción. Por ejemplo, se pueden solucionar problemas al generar las variedades morfológicas propias de la LM, mediante *machine learning* o la introducción de diccionarios (España-Bonet y Costa-Jussa, 2016: 10). Otros enfoques han sido, por ejemplo, herramientas automáticas basadas en reglas capaces de poseer resultados de traducción de sistemas estadísticos, es decir, herramientas que incorporen una serie de reglas capaces de corregir los resultados.

Existen también modelos cuya raíz se encuentra en los sistemas basados en reglas, pero que cuentan con características de los sistemas estadísticos, aunque los modelos basados en reglas, como ya se advirtió en anteriores apartados, requieren una mayor inversión de tiempo y mano

de obra, ya que tanto las reglas como los diccionarios se recopilan y escriben manualmente. No obstante, según España-Bonet y Costa-Jussa (2016: 10), si se dispone de unas reglas de calidad, estos sistemas producen mejores traducciones desde el punto de vista sintáctico y hacen un mejor trabajo al reordenar oraciones. En este caso, el beneficio que aportan los sistemas basados en corpus (TAE y TA basada en ejemplos) a los sistemas basados en reglas es una más acertada selección léxica y una mayor fluidez (España-Bonet y Costa-Jussa, 2016: 10; Sánchez-Martínez, 2008: 06).

Según Oliver (2014: 20), existen tres enfoques diferentes al combinar los sistemas y formar un sistema híbrido:

1. **Combinación de varios sistemas de TA (*multi-engine approach*):** la traducción se realiza con varios sistemas en paralelo y el texto final se genera a partir de los resultados de cada uno de estos sistemas.
2. **Generación estadística de reglas:** se trata del uso de datos estadísticos para crear reglas léxicas y sintácticas. Una vez se han creado estas normas lingüísticas a partir de datos estadísticos, el sistema funcionará como un sistema basado en reglas. Así se evita tener que crear estas reglas de forma manual.
3. **Multi-etapas (*multi-pass approach*):** en este enfoque se procesa el texto de entrada varias veces en serie. Por ejemplo, primero un sistema basado en reglas procesa el texto de entrada y, a continuación, un sistema estadístico realiza el procesamiento del texto de salida del sistema basado en reglas, el cual produce la traducción final. De esta forma, los sistemas estadísticos no tienen que procesar tanta información y el sistema basado en reglas no tendrá que ser un sistema completo, (lo que reduce el trabajo manual).

Aunque gran parte de los estudios sobre sistemas híbridos se basen en investigar la combinación de sistemas de TAE y de TA basada en reglas, también surgieron otros sistemas que combinaban, por ejemplo, TA basada en reglas con TA basada en ejemplos, TAE con TA basada en ejemplos (donde la base es basada en ejemplos, pero luego se utiliza un modelo lingüístico estadístico en el proceso de recombinación del texto) e incluso sistemas de TA basada en reglas y TAE con TA basada en ejemplos (donde la base es una combinación de un sistema estadístico y reglas lingüísticas, pero se utilizan plantillas de traducción, propias de la TA basada en ejemplos).

Se podría afirmar que los sistemas híbridos despertaron un gran interés en el ámbito de la investigación por su potencial y su practicidad, no obstante, a partir de 2016, aparece un nuevo paradigma que da un giro a la forma en que la TA había funcionado hasta ahora, la TA neuronal por lo que poco a poco han ido apareciendo también estudios sobre sistemas híbridos que

incluyen características de los sistemas de TAN. En el siguiente apartado se tratarán en profundidad las características de la TAN y se observará su evolución desde su aparición hasta la actualidad.

#### 2.3.2.4. *TA Neuronal*

Durante años, el paradigma más novedoso dentro de la TA fue el de los sistemas de TAE. De manera más reciente, los sistemas de TAN se han convertido en el centro de atención tanto en el ámbito de la investigación como en la industria, llegando a desbancar a su predecesora, la TAE. Según Castilho et al. (2017: 109) esto es en parte por los buenos resultados que han obtenido las traducciones llevadas a cabo con estos sistemas en las evaluaciones automáticas como BLEU. Además, este modelo es capaz de suplir muchas de las carencias de los sistemas anteriores de TA (Wu et al., 2016: 02).

Algunos de los avances que se observarán a continuación se llevaron a cabo con el ánimo de conseguir sistemas multilingües que tradujeran de manera fiable sin realizar demasiados cambios al algoritmo del sistema principal. De esta manera, el enfoque de la TAN más novedosa es de tipo aprendizaje *end-to-end* (de extremo a extremo), es decir, que agrupa en una misma red neuronal las diferentes etapas del entrenamiento, en vez de dividirlo en diferentes. Esto es especialmente beneficioso para planteamientos de sistemas de TA multilingües (Johnson et al., 2017: 340). En los párrafos siguientes, se ofrece una visión global de cómo funcionan estos sistemas, de las definiciones que se han proporcionado a lo largo de los años, de los principales avances y aplicaciones de la TAN, así como sus ventajas e inconvenientes y las diferencias con la TAE, entre otros aspectos.

La relativamente reciente propuesta de la TAN está basada en redes neuronales artificiales (RNA) y aprendizaje automático (ML o *machine learning* en inglés) y es actualmente uno de los paradigmas más punteros de TA, tanto en el ámbito de la investigación como en el de la industria. Esta ha mostrado gran potencial en sus resultados si se compara con la TAE, incluso en tareas a gran escala, hasta el punto en que se ha llegado a afirmar que sustituirá por completo a los sistemas estadísticos (Aharoni, Johnson y Firat, 2019: 3874; Casacuberta y Peris-Abril, 2017: 68; Jean, Cho, Memisevic y Bengio, 2014; Schulz, Aziz y Cohn, 2018; Sennrich, Haddow y Birch, 2015).

Según Forcada (2017: 292), se podría definir la TAN de la siguiente manera:

Neural machine translation is a new breed of corpus-based machine translation (also called data-driven or, less often, corpus-driven machine translation). It is trained on huge corpora of pairs of source language segments (usually sentences) and their

translations, that is, basically from huge translation memories containing hundreds of thousands or even millions of translation units. In this sense, it is similar to the statistical machine translation technology that was the state of the art until very recently, but uses a completely different computational approach: neural networks.

Casacuberta y Peris-Abril (2017: 66) indican que un modelo de TAN es un tipo de modelo estadístico compuesto por unidades de procesamiento interconectadas. Los parámetros para estos modelos se estiman a partir de corpus paralelos (al igual que en los modelos basados en corpus, como el estadístico) mediante algoritmos de aprendizaje automático (ML o *machine learning* en inglés) y potentes procesadores gráficos. En estos sistemas de TA, en los que se aplican los modelos neuronales, las palabras se representan mediante vectores y para procesar las oraciones se suelen utilizar redes neuronales recurrentes (RNN o *Recurrent Neural Networks* en inglés), en las cuales las neuronas se retroalimentan con sus propias salidas de forma directa o indirecta (Casacuberta y Peris-Abril, 2017: 66).

El nombre de «traducción automática neuronal» proviene precisamente de las redes neuronales artificiales en las que se basa la TAN. Dichas redes están compuestas de miles de unidades artificiales que simulan neuronas, en el sentido de que estas se activan (se excitan o se inhiben) en función del estímulo que reciban de otras neuronas y de la fuerza de las conexiones entre neuronas a través de las cuales fluyen estos estímulos (Forcada, 2017: 292).

La TAN está estrechamente relacionada con el aprendizaje automático y, desde el punto de vista de esta disciplina, la traducción se puede entender como el aprendizaje de la distribución condicional  $p(y | x)$ , de una oración en LM ( $y$ ), dada una oración en LO ( $x$ ). Las redes neuronales aprenden dicha distribución condicional a partir de un corpus paralelo bilingüe (Pouget-Abadie, Cho, Marriënboer, Bahdanau y Bengio, 2014). En otras palabras, la TAN se basa en una red neuronal que lee la oración en LO y otra que genera la traducción en LM. El sistema se entrena, por lo tanto, de manera conjunta para maximizar la probabilidad de obtener la traducción correcta de determinada oración en LO, mediante el uso de corpus bilingües alineados, sin necesidad de información lingüística externa (Castilho et al., 2017: 110; Jean, Cho, Memisevic y Bengio, 2014; Luong, Pham y Manning, 2015).

De entre las principales características de la TAN, se encuentra el hecho de que los modelos multilingües de TAN son capaces de traducir entre diferentes LO y LM dentro de un mismo modelo simple que soporta la traducción entre múltiples pares de lenguas (Arivazhagan et al., 2019). En este tipo de modelos, las palabras y las frases se representan de forma numérica a

partir de vectores, lo que también ha permitido la implementación de las redes neuronales (Casacuberta y Peris-Abril, 2017: 68).

Estos modelos basados en redes neuronales suelen estar formados por un codificador (*encoder*) y un decodificador (*decoder*). Tienen un funcionamiento muy complejo ya que, por un lado, hay que comprender cómo se entrenan estos sistemas y, por otro, conocer algunas de las arquitecturas más típicas de este tipo de modelos. A continuación, se ofrece una perspectiva general de estos aspectos, para que su estudio sea más accesible a lingüistas y a estudiantes por igual.

Hay que tener en cuenta que los procesos de entrenamiento y traducción son muy costosos computacionalmente, por lo que se necesitan procesadores gráficos (GPU) muy potentes, los cuales realizan operaciones a gran velocidad. Además, los desarrolladores disponen de herramientas *software* para el entrenamiento de dichos modelos de TAN, como, por ejemplo, TensorFlow, GroundHog o Keras, entre otras (Casacuberta y Peris-Abril, 2017: 70).

El potencial de la TAN se encuentra en la posibilidad de que esta aprenda directamente, siguiendo el principio del *end-to-end*, las correlaciones entre el texto de entrada y el de salida. Este tipo de TA tiene una arquitectura formada normalmente por dos redes neuronales recurrentes (RNN o *recurrent neural networks* en inglés), una de las cuales recibe el texto de entrada y la otra genera el texto traducido. Más adelante se quiso mejorar el modelo neuronal, por lo que se incorporó un mecanismo de atención, el cual ayuda a procesar oraciones de entrada más largas (Cho, Marriënboer, Bahdanau y Bengio, 2014; Peris-Abril, Domingo-Ballester y Casacuberta, 2017: 02; Wu et al., 2016: 01).

Este tipo de sistemas está basado en corpus paralelos, al igual que los modelos anteriores como el estadístico (Sennrich, Haddow y Birch, 2015), los cuales se crearán antes del entrenamiento. Como explica Forcada (2017: 294) la TAN se relaciona con el aprendizaje profundo o *deep learning*, de manera que las representaciones se realizan por capas, las cuales suelen contener cientos de unidades neuronales. Estas unidades están unidas por pesos (*weights*), los cuales conectan todas las unidades de una capa con las unidades de la siguiente. Según el peso y la magnitud de la conexión entre diferentes neuronas, se determinará el comportamiento de la red neuronal.

Estos pesos pueden ser positivos o negativos, de manera que, si el estímulo se recibe a través de una conexión con peso positivo, la neurona tenderá a excitar la neurona con la que esté conectada. Por el contrario, si dicho estímulo es negativo, la neurona tenderá a inhibir la neurona con la que está conectada. El entrenamiento consiste en determinar el valor de los

diferentes pesos de las conexiones neuronales, a partir de un corpus de entrenamiento (Casacuberta y Peris-Abril, 2017: 69), para obtener los resultados deseados. Una vez hecho esto, una función de activación asigna valores de  $y$  a valores de  $x$ . Las activaciones de cada neurona en cada capa se conocen como «vectores» y están formados por valores (por ejemplo, +0.3, 0, -0.23, ...), a partir de los cuales se llevará a cabo la traducción. (Forcada, 2017: 293-295).

En TAN, las palabras o las unidades más pequeñas (caracteres o secuencias cortas de caracteres) se procesan en paralelo por las neuronas de cada capa. Los estados de activación de cada neurona dentro de grandes grupos neuronales se entrenan para crear representaciones de las palabras y sus contextos. A estas representaciones se les suele llamar *embeddings* y se aprenden automáticamente a partir de textos monolingües mediante redes neuronales especializadas que aprenden a reproducir una palabra específica en un contexto concreto, a partir de algunas de las palabras que aparecen a su derecha y a su izquierda. De este modo, el acto de traducir en TAN implica codificar para decodificar (Forcada, 2017: 296).

Si se profundiza en cuestiones estadísticas, un sistema de TAN es una red neuronal que calcula la probabilidad de que una oración en LM sea la traducción de una oración en LO, es decir,  $p(y | x)$  (Castilho et al., 2017: 110; Jean, Cho, Memisevic y Bengio, 2014; Kalchbrenner et al., 2016; Luong, Pham y Manning, 2015). Wu et al. (2016: 03) descomponen dicha formulación en la Ecuación 1 siguiente:

$$P(X) = P(Y|(x_1, x_2, x_3 \dots, x_m))$$

$$= \prod_{i=1}^N P(y_i | y_0, y_1, y_2, \dots, y_{i-1}; x_0, x_1, x_2, \dots, x_m)$$

*Ecuación 1. Representación de la probabilidad de que una oración en LM sea la traducción de una oración en LO. Extraído de Wu et al. (2016: 3).*

Según dicha ecuación,  $X = x_1, x_2, x_3 \dots, x_m$  representa la secuencia de  $M$  símbolos en la oración en LO e  $Y = y_1, y_2, y_3 \dots, y_n$  representa la secuencia de  $N$  símbolos en la oración en LM (Wu et al., 2016: 03). Kalchbrenner et al. (2016), lo expresa de manera similar, donde  $t$  pasa a ser la oración en LM y  $s$  la oración en LO (véase Ecuación 2):

$$p(t|s) \prod_{i=0}^N p(t_i | t_{<i}, s)$$

*Ecuación 2. Representación de la probabilidad de que una oración en LM sea la traducción de una oración en LO. Extraído de Kalchbrenner et al. (2016).*

Como es posible ver en la Ilustración 7, a continuación, primero el codificador combina una oración vacía, que representa como «E(' ')» e indica el comienzo de la oración, con la primera palabra de la oración en LO, representada como «e('my')». Así, el codificador crea la codificación

«E('My')». Luego sigue uniendo representaciones de las palabras de la oración en LO, de manera que combina «E('My')» con «e('flight')» para obtener «E('My flight')». Este proceso se sigue hasta haber codificado la oración entera, «E('My flight is delayed.')». A continuación, otro codificador procesará la oración, en la dirección opuesta, de derecha a izquierda, «E('delayed is flight My')».

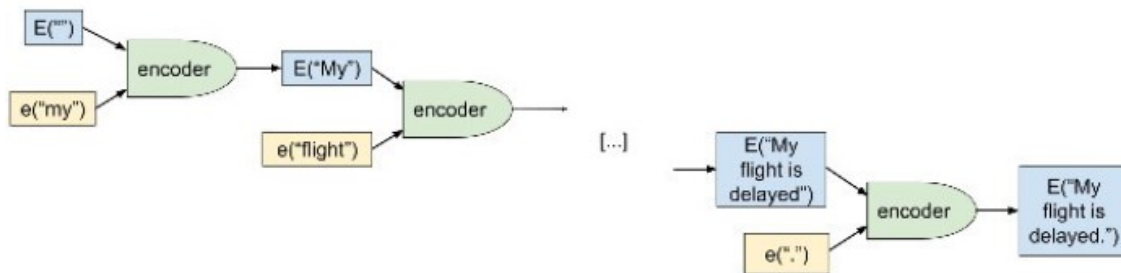


Ilustración 7. Codificación de la oración "My flight is delayed". Extraída de Forcada (2017).

Tras haber codificado la oración en LO entera, llega el paso de la decodificación. En este momento el decodificador generará una traducción condicionada por la oración original, a partir del vector que haya generado el codificador (Bahdanau, Cho y Bengio, 2014; Casacuberta y Peris-Abril, 2017: 69; Lakew et al., 2018). El decodificador, que es otra red neuronal recurrente, genera una palabra en LM en función de la palabra generada anteriormente, el estado de la red neuronal en el instante anterior y una representación de la frase origen obtenida a partir del modelo de atención y de la información proporcionada por el codificador (Casacuberta y Peris-Abril, 2017: 69).

El decodificador ofrece la probabilidad de que cierta palabra en LM sea la continuación de la oración en LM que ya se había ido generando anteriormente, es decir, que la traducción se genera mediante la elección de la palabra más probable para cada posición de la oración (Forcada, 2017: 295). En la Ilustración 8 se observa cómo funcionan los decodificadores:

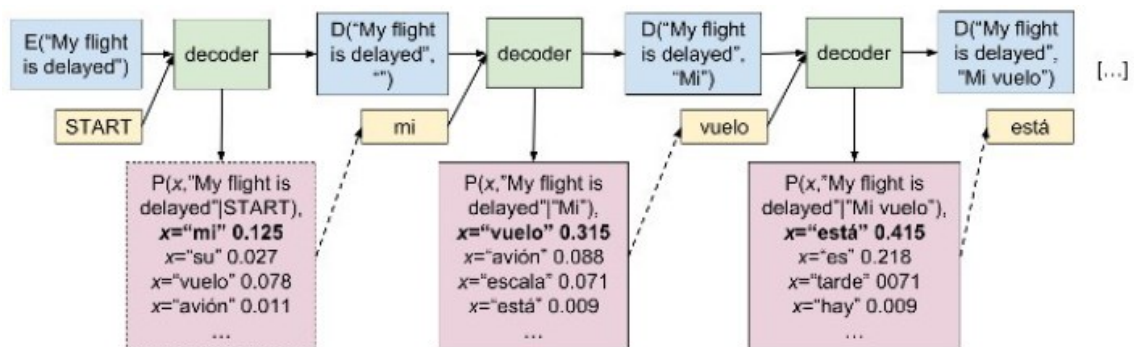


Ilustración 8. Decodificación de inglés a español de la oración "My flight is delayed". Extraída de Forcada (2017).



Siguiendo el esquema de la Ilustración 8, extraída de Forcada (2017), el proceso de decodificación parte de la oración completa que el codificador había codificado en el proceso previo, «E('My flight is delayed.')». El decodificador creará dos vectores, uno con el estado inicial «D('My flight is delayed., '))», donde " representa una secuencia vacía de palabras; el otro vector será un vector de probabilidades con todas las posibles palabras  $x$  que puedan ir al principio de la oración, es decir, « $p(x| \text{'My flight is delayed., '})$ ». Entonces el decodificador asignará a  $x$  la probabilidad de que su traducción en español sea «Mi» ( $x= \text{'Mi'}$ ).

A continuación, el decodificador lee «D('My flight is delayed., '))» y la palabra «Mi» y generará dos vectores. Por un lado, el siguiente estado del decodificador «D('My flight is delayed., 'Mi')» y un vector de probabilidades para todas las posibles palabras  $x$  que puedan ir en la segunda posición de la oración « $p(x| \text{'My flight is delayed., 'Mi'})$ ». El decodificador debería asignar a  $x$  la probabilidad de que su traducción en español sea «vuelo» ( $x= \text{'vuelo'}$ ). Este proceso se repetirá sucesivamente hasta que se haya decodificado la oración por completo.

Para llevar a cabo este proceso, algunos sistemas utilizan redes neuronales recurrentes que permiten determinar el contexto de cierta palabra teniendo en cuenta las palabras que aparecen delante de ella y las que aparecen detrás de ella, lo que permite al sistema tomar decisiones informadas (Peris-Abril y Casacuberta, 2015: 110).

La arquitectura más común de entre los sistemas de TAN, es el modelo codificador-decodificador o *encoder-decoder* (Bahdanau, Cho y Bengio, 2014; Jean, Cho, Memisevic y Bengio, 2014). La red neuronal estima una traducción a LM condicionada por la oración de entrada en LO. Dicha red neuronal artificial está compuesta por una red para LO (codificador), la cual codifica la secuencia en LO en una representación numérica, un vector, y una red para LM (decodificador), el cual utiliza esta representación generada por el codificador para producir una oración traducida (Bahdanau, Cho y Bengio, 2014; Kalchbrenner et al., 2016).

El sistema codificador-decodificador se entrena de manera conjunta para maximizar la probabilidad de conseguir una traducción correcta, dada una oración en LO concreta (Bahdanau, Cho y Bengio, 2014). El desempeño de los sistemas con esta arquitectura codificador-decodificador formada por redes neuronales recurrentes se mejoró mediante un mecanismo de atención, el cual alinea las unidades de traducción en LM con las de LO y ayuda a procesar oraciones de entrada más largas (Bentivogli et al., 2016; Britz et al., 2017: 01; Cho, Marriënboer, Bahdanau y Bengio, 2014; Peris-Abril, Domingo-Ballester y Casacuberta, 2017: 02; Wu et al., 2016: 01).

La red neuronal que forma el codificador lee y codifica una oración en un vector de longitud fija (Bahdanau, Cho y Bengio, 2014; Lakew et al., 2018). De la misma forma, según Casacuberta y Peris-Abril (2017: 69), el codificador es una red neuronal recurrente que «analiza de izquierda a derecha y de derecha a izquierda la frase origen para producir una representación vectorial de la misma». Las capas que forman los codificadores se organizan en celdas o *gating structures* con capacidad de aprender a olvidar textos de entrada que no son relevantes y de aprender a recordar aquellos que sí lo son. Las estructuras más comunes son las memorias LSTM (*Long-short term memories* en inglés) y las unidades recurrentes cerradas (GRU o *gated recurrent units* en inglés) (Forcada, 2017: 297). En la Ilustración 9, se observan las estructuras, según Tan et al. (2020: 05-06) de las RNN típicas y de las BRNN (bidirectional recurrent neural networks) utilizadas en el modelo codificador-decodificador original de Bahdanau, Cho y Bengio (2014):

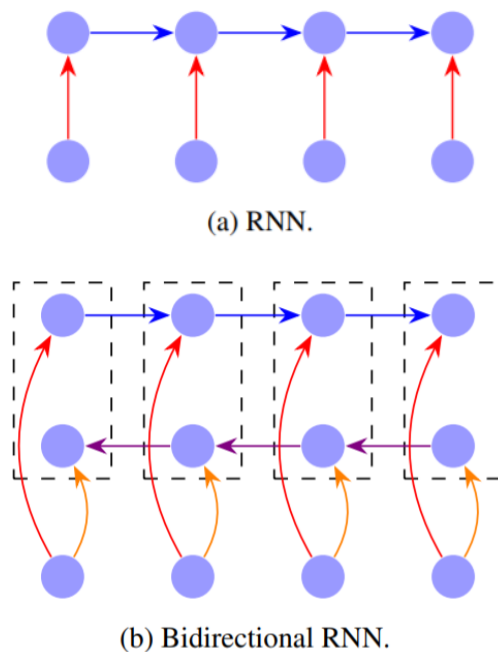


Ilustración 9. Estructuras de una RNN y de una BRNN.

Como se ha podido observar, la estructura más común es la de codificador-decodificador con un mecanismo de atención y estos sistemas se entrenan mediante un corpus paralelo. Según Bahdanau, Cho y Bengio (2014: 02):

In neural machine translation, we fit a parameterized model to maximize the conditional probability of sentence pairs using a parallel training corpus. Once the conditional distribution is learned by a translation model, given a source sentence a corresponding translation can be generated by searching for the sentence that maximizes the conditional probability.

Los modelos basados en el modelo clásico codificador-decodificador se entrenan de manera conjunta, mediante un corpus paralelo, de extremo a extremo para maximizar la probabilidad de que las traducciones sean correctas. El codificador es una red neuronal recurrente bidireccional que codifica una oración de origen,  $x = (x_1, \dots, x_m)$ , en una secuencia de vectores de estado o *hidden state vectors* que va en dos direcciones:  $(\vec{h}, \dots, \vec{h}_m)$  y  $(h^{\leftarrow}, \dots, h^{\leftarrow}_m)$ . El decodificador, que es otra red neuronal recurrente, utiliza estas representaciones numéricas para predecir la secuencia en LM  $y = (y_1, \dots, y_n)$ , y el mecanismo de atención se utiliza para mejorar la calidad de la traducción, ya que sugiere qué parte de la oración en LO ayuda en mayor manera en el proceso de predicción. Cada oración  $y_i$  se origina basándose en un estado anterior  $s_i$ , la palabra traducida anteriormente,  $y_{i-1}$ , y un vector de contexto,  $c_i$  (Cho et al., 2014; Lakew, et al, 2018; Sennrich, Haddow y Birch, 2015).

Esto se observa en el modelo presentado por Peris-Abril, Domingo-Ballester y Casacuberta (2017) (Ilustración 10), basado en el modelo original de Bahdanau, Cho y Bengio (2014):

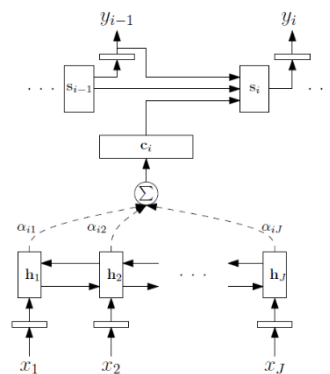


Ilustración 10. Modelo de TAN basado en codificador-decodificador. Extraído de Peris-Abril, Domingo-Ballester y Casacuberta (2017: 6), siguiendo el modelo de Bahdanau, Cho y Bengio, 2014.

Peris-Abril, Domingo-Ballester y Casacuberta (2017: 06), explican su versión del modelo de la siguiente manera:

The bidirectional RNN encoder processes the source sentence  $x_1^J$  and generates the sequence of hidden states  $h_1^J$ . The alignment model assigns a weight  $\alpha_{ij}$  to each hidden state, according to the previous state of the decoder  $s_{i-1}$  and the hidden state from the encoder. The states from the encoder are then combined in order to obtain a context vector  $c_i$ . The decoder generates the next word  $y_i$  according to the context vector  $c_i$ , its previous hidden state  $s_{i-1}$  and the previously generated word  $y_{i-1}$ .

Aunque la arquitectura más utilizada para desarrollar los modelos de TAN sea la de codificador-decodificador con redes neuronales recurrentes, posteriormente se han ido desarrollando

sistemas neuronales cada vez más complejos y se han introducido, avances como las redes neuronales convolucionales (CNN o *Convolutional Neural Networks* en inglés) o la traducción *zero-shot*.

Cho, Marriënboer, Bahdanau y Bengio (2014) hablan de dos enfoques diferentes sobre la TAN. Por un lado, la basada en redes neuronales recurrentes y, por otro, la basada en redes neuronales convolucionales. Algunos de los enfoques más recientes ya no usan las típicas arquitecturas recurrentes de tipo codificador-decodificador como las que se han explicado anteriormente. En el caso de las arquitecturas basadas en redes neuronales convolucionales, en vez de generar una codificación de la oración original entera a partir de representaciones de palabras sueltas en LO, el decodificador produce representaciones de cada palabra teniendo en cuenta el conjunto de palabras a su alrededor. Por ejemplo, puede tener en cuenta aproximadamente dos o tres palabras a la izquierda o a la derecha de la palabra en cuestión. Citando el ejemplo de Forcada (2017: 298), estas redes actuarían de la siguiente manera:

For instance, our sentence 'My flight is delayed', conveniently padded to form 'NULL NULL My flight is delayed. NULL NULL' is turned into a series of context-informed representations  $R(\text{'NULL NULL My flight is'})$ ,  $R(\text{'NULL My flight is delayed'})$ ,  $R(\text{'My flight is delayed .'})$ ,  $R(\text{'flight is delayed . NULL'})$ , and  $R(\text{'is delayed . NULL NULL'})$ ; then, these representations are taken again in groups of 5 and used to generate a series of deeper representations; this is repeated (convoluted) a couple of times. Then a similar scheme is used to generate representations of output words.

Es decir, que se calculan vectores para los bi-gramas, tri-gramas o más tipos de n-gramas de la oración. Se puede observar cómo funcionan estas redes neuronales convolucionales en la Ilustración 11:

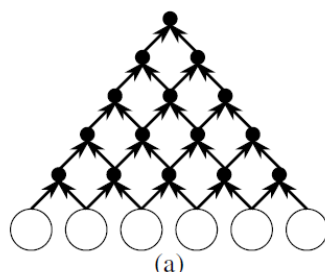


Ilustración 11. Representación gráfica de una red neuronal convolucional. Extraído de Cho, Marriënboer, Bahdanau y Bengio, 2014.

En aquellos sistemas que utilizan redes neuronales convolucionales, la información de cada nivel se comparte en la red neuronal completa, por lo que cada filtro en la convolución tiene una

visión de la oración en LO completa, de la cual la red seleccionará algunas características. De esta manera, se podrán formar codificaciones que consideren toda la información de la oración durante el proceso de codificado (Cho, Marriënboer, Bahdanau y Bengio, 2014; Lamb y Xie, 2016: 01). Gehring et al. (2016) aseguran que este tipo de modelos tienen arquitecturas más simples y son más rápidos, ya que las capas convolucionales facilitan la identificación de las relaciones entre elementos de una oración.

Otro de los avances mencionados anteriormente es la traducción automática *zero-shot*. El principio de la traducción *zero-shot* se basa en los sistemas capaces de traducir dentro de un mismo modelo multilingüe entre pares de lenguas con las que no han estado en contacto durante su entrenamiento (Aharoni, Johnson y Firat, 2019: 3874; Al-shedivat y Parikh, 2019: 1184; Johnson et al., 2017: 341; Lakew et al., 2018). Gracias al aprendizaje *zero-shot*, se reduce el trabajo y la supervisión manual (Dinu, Lazaridou y Baroni, 2015: 01). Un ejemplo sería un sistema multilingüe entrenado con corpus paralelos para los pares PT-EN y EN-ES. Si el objetivo fuera una traducción de PT→ES, en este caso, el sistema no dispone de corpus de PT-ES, por lo que traduciría primero de PT→EN y, a continuación, de EN→ES, sin que durante el entrenamiento se le haya enseñado a traducir de PT→ES.

Otros avances incluyen la mejora de los mecanismos de atención mediante la inclusión de más información lingüística y más lenguas a los modelos (Castilho et al., 2017: 110). Además, recientemente, se han planteado enfoques para modelos que permitan entrenar sistemas multilingües que traduzcan entre diversos pares de lenguas dentro de un único modelo (Aharoni, Johnson y Firat, 2019: 3874; Siddhant et al., 2019). Es decir, que dentro de un mismo sistema de TA multilingüe no tengan que existir modelos distintos para cada par de lenguas. Por último, también se ha mejorado la calidad de las traducciones a lenguas con menos recursos (Siddhant et al., 2019).

#### 2.3.2.5. Traducción Neuronal de Google

Como se ha observado, a lo largo de los años, se han desarrollado multitud de sistemas y modelos con mejoras para conseguir traducciones más precisas. No obstante, entre los avances relativos a TAN, es vital mencionar el gran cambio que experimentó el Traductor de Google en 2016 al pasar de ser un sistema estadístico a un sistema basado en TAN.

Schuster, Johnson y Thorat (2016) definen el sistema de traducción automática neuronal de Google (GNMT o *Google Neural Machine Translation* en inglés), como un marco de aprendizaje *end-to-end* entrenado a partir de millones de ejemplos, que introdujo una sustancial mejora en la calidad de los resultados de traducción con respecto a sistemas anteriores. Este sistema

multilingüe es capaz de traducir entre varios pares de lenguas dentro de un único modelo sin que se deban realizar grandes cambios en la arquitectura del modelo. Esto es posible gracias a la introducción de una etiqueta, que Schuster, Johnson y Thorat (2016) denominan *token*, que indica el idioma hacia el cual se va a traducir la oración en LO. Este *token* o etiqueta tendría el siguiente aspecto «<2es> Hello, how are you? -> Hola, ¿cómo estás?». En el ejemplo anterior, se muestra cómo se ha añadido la etiqueta «<2es>» para indicar que el texto se debe traducir a español.

En anteriores sistemas, no era posible tener un único modelo dentro de un mismo sistema multilingüe. Las únicas opciones eran tener un modelo para cada par de lenguas, tener un codificador común cuando la LO era la misma (p.ej.: EN-ES, EN-FR, EN-NL) o tener un decodificador común cuando la LM era la misma (p.ej.: ES-EN, FR-EN, NL-EN). Gracias a esta simplificación del modelo, es mucho más sencillo ampliar el sistema mediante la incorporación de nuevas lenguas. Esto también beneficia a la traducción entre pares de lenguas con menos recursos, puesto que se comparten recursos entre lenguas en un mismo sistema.

En su artículo, *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*, uno de los estudios de referencia en este tema, Wu et al. (2016: 02) aseguran que GNMT ha conseguido reducir los errores de traducción hasta en un 60 % en comparación con sistemas anteriores para aquellos pares de lenguas con un mayor uso, como inglés-francés, inglés-español o inglés-chino. Wu et al. (2016: 04) también publican en su artículo la arquitectura del modelo, el cual se muestra en la Ilustración 12:

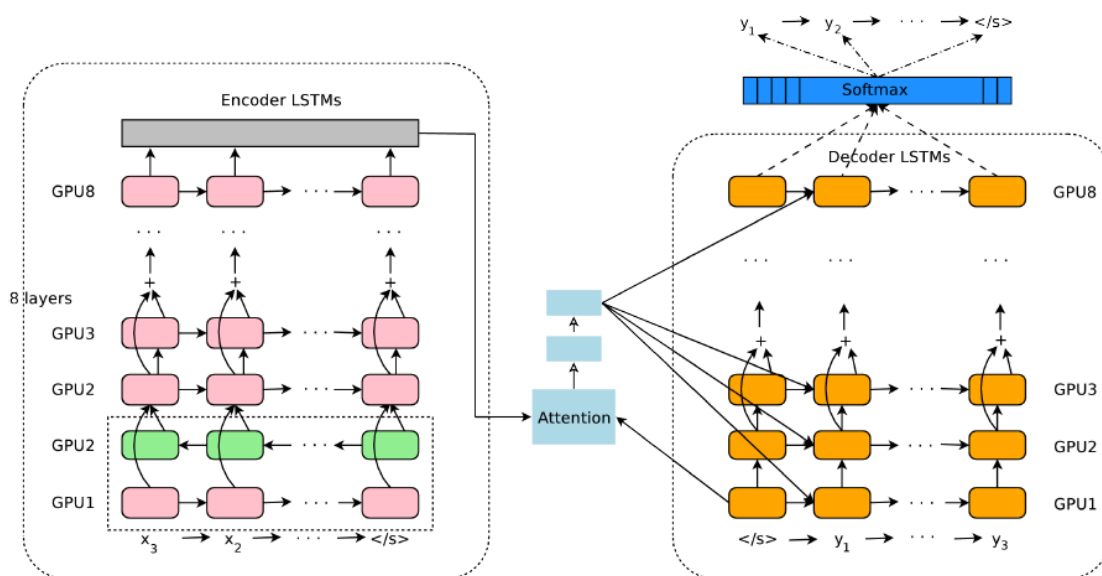


Ilustración 12. Modelo de TAN de Google. Extraído de Wu et al. (2016: 4).

Este modelo también consta de una arquitectura codificador (izquierda) decodificador (derecha) e incorpora además un módulo de atención (centro). La parte inferior del codificador es una capa bidireccional, cuyos nodos recopilan información de derecha a izquierda y viceversa. Las otras capas del codificador son unidireccionales. En total hay 8 capas en el codificador y 8 capas en el decodificador. El modelo está dividido en múltiples unidades de procesamiento gráfico (GPU o *graphics processing unit*, en inglés) para acelerar el entrenamiento. Mientras tanto el módulo de atención obtendrá información relevante sobre el contexto, la cual pasará al decodificador y le permitirá centrarse en diferentes partes de la oración original durante el proceso de decodificación (Wu et al., 2016: 03-04).

### 2.3.2.6. *Transformer*

Si bien la TAN de Google significó una gran innovación, durante los últimos años, se ha desarrollado un nuevo modelo llamado *Transformer*. Según Vaswani et al. (2017: 01), este modelo cuenta con una arquitectura más sencilla basada solamente en mecanismos de atención y que prescinde totalmente de las redes recurrentes y convolucionales. Además, Vaswani et al. (2017: 01) sugieren que los resultados de traducción producidos por el *Transformer* son de mayor calidad que los que ofrecen modelos anteriores y que este modelo es más fácil de entrenar.

Como los modelos anteriores, el *Transformer* sigue la estructura codificador-decodificador (Ilustración 13):

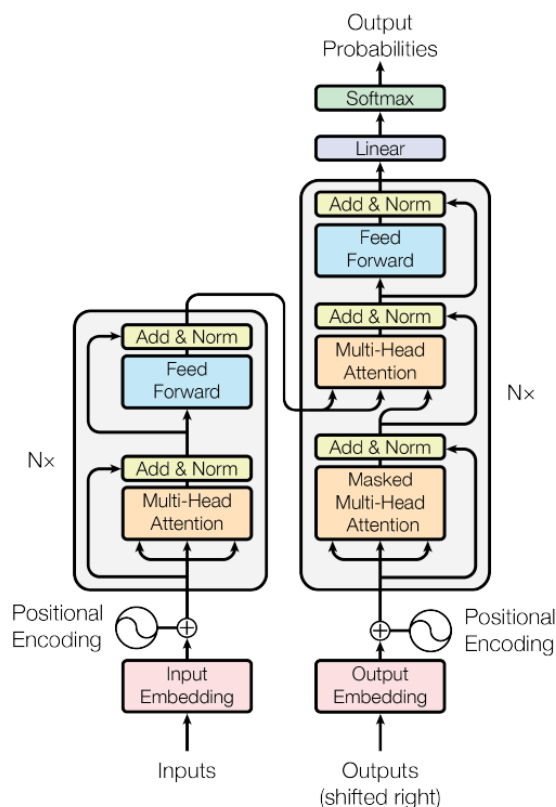


Ilustración 13. Estructura del modelo Transformer según Vaswani et al. (2017: 3).

Como se puede observar en la Ilustración 13, el codificador está compuesto por seis capas idénticas conectadas, cada una de las cuales está a su vez compuesta por dos capas más. La primera capa es un mecanismo multicabezal de atención (*Multi-Head Attention Mechanism*, en inglés) y la segunda es una red neuronal prealimentada (*Feedforward Neural Network*, en inglés) completamente conectada. Además, se añade una conexión residual sobre cada una de las capas, que irá seguida de una capa de normalización. Por otro lado, el decodificador está compuesto por otras seis capas idénticas conectadas con una estructura similar a la del codificador. En este caso, además de las dos subcapas con que también cuenta el codificador, se añade una tercera capa multicabezal de atención que actuará sobre los resultados del codificador (Vaswani et al., 2017: 03).

Tras haber detallado el funcionamiento de este novedoso enfoque, en el apartado siguiente, se han detallado las principales ventajas e inconvenientes que presenta la traducción automática neuronal con respecto a modelos anteriores.

### 2.3.2.7. Ventajas e inconvenientes de la TAN

Como se ha podido observar en secciones anteriores, la TAN ha supuesto un gran avance en los ámbitos de la traducción automática, del procesamiento del lenguaje natural y de la lingüística computacional. Esta ha supuesto una gran mejora con respecto a modelos anteriores de TA,



como la TAE, la cual fue el paradigma dominante hasta la aparición de la TAN. A continuación, se ha proporcionado una visión general de algunas de las ventajas e inconvenientes de la TAN.

De entre las principales ventajas que ofrece la TAN se encuentra la posibilidad de entrenar un único modelo a partir del cual los sistemas de TAN multilingüe puedan traducir desde múltiples LO, a varias LM (Aharoni, Johnson y Firat, 2019: 3874). Además, los sistemas de TAN se entrenan de forma *end-to-end*, lo que simplifica mucho la arquitectura de dichos modelos con respecto a sistemas anteriores de TA (Bentivogli et al., 2016). De esta manera, no hace falta que estos contengan tablas enormes con oraciones de ejemplo, ni modelos lingüísticos, lo que significa que los sistemas de TAN ocupan mucha menos memoria que sus predecesores, los sistemas de TAE (Cho, Marriënboer, Badahnau y Bengio, 2014; Luong, Pham y Manning, 2015).

De la misma forma, Jean, Cho, Memisevich y Bengio (2014), aseguran que la TAN hace un uso mucho menor de la memoria que la TAE, ya que esta última se sirve de grandes tablas con pares de oraciones. Es más, al contrario que los sistemas anteriores de TA, en que los componentes y las funciones se debían entrenar por separado, los sistemas de TAN entrenan todos sus componentes de manera conjunta, lo que da como resultado un mejor desempeño en los resultados de traducción (Jean, Cho, Memisevich y Bengio, 2014). Otra de las mejoras que introduce la TAN es la reducción del tamaño del vocabulario para mejorar la eficiencia y acortar los tiempos de entrenamiento (Sennrich, Haddow y Birch, 2015).

La TAN además ha conseguido muy buenos resultados en las evaluaciones en comparación con los que había obtenido anteriormente la TAE, sobre todo en los pares de lenguas más comunes y en aquellas lenguas ricas morfológicamente (Casacuberta y Peris-Abril, 2017: 71; Castilho, 2017: 110; Sánchez-Gijón, Moorkens y Way, 2019: 31; Wu et al., 2016). En palabras de Castilho (2017: 109):

More recently, the neural approach (NMT) has emerged as a new paradigm in MT systems, raising interest in academia and industry by outperforming phrase-based statistical systems (PBSMT), based largely on impressive results in automatic evaluation.

Los resultados de traducción de los sistemas de TAN cuentan con mayor fluidez y naturalidad que los resultados conseguidos con sistemas estadísticos (Sánchez-Gijón, Moorkens y Way, 2019: 31). Además, el aprendizaje por transferencia ha permitido que aquellas lenguas con menos recursos se entrenen con lenguas con más recursos, por lo que la calidad de la traducción para las lenguas con menos recursos mejora de manera sustancial (Aharoni, Johnson y Firat, 2019: 3874).

En referencia al esfuerzo de posesición de contenidos traducidos mediante sistemas de TAN, se ha observado que este se ha visto reducido considerablemente. Bentivogli et al. (2016) indican que dicho esfuerzo ha disminuido en un 26 % en comparación con los sistemas de TAE, en cuanto a errores léxicos, morfológicos y sintácticos, y asegura que la TAN «has pushed ahead the state-of-the-art», sobre todo para idiomas ricos morfológicamente. Sin embargo, en esta afirmación, Bentivogli et al. (2016) no tuvieron en cuenta el tiempo de posesición (Forcada, 2017: 304).

Vistas las ventajas que la TAN presenta con respecto a sistemas anteriores de TA, toca ahora echar un vistazo a los posibles problemas e inconvenientes que se desprenden de esta. Sánchez-Gijón, Moorkens y Way (2019: 31) aseguran que, pese a que la TAN ha superado según las métricas a otros sistemas de TA, esta se sigue enfrentando a algunos desafíos cuando se utiliza en entornos profesionales.

En párrafos anteriores, se han especificado las ventajas que presenta el entrenamiento *end-to-end* que se lleva a cabo los sistemas de TAN, ya que este tipo de entrenamiento permite solventar algunos de los puntos débiles que presentaban los sistemas anteriores. No obstante, los sistemas neuronales son costosos de entrenar desde el punto de vista computacional y, aunque hayan ofrecido buenos resultados con sistemas robustos y suficiente potencia computacional, estos sistemas cometen errores si las oraciones originales de entrada contienen palabras desconocidas (Ha, Niehues y Waibel, 2017; Wu et al., 2016: 01).

Para superar estos problemas relacionados con la capacidad computacional necesaria para que estos sistemas funcionen correctamente, se necesita un *hardware* muy potente. No solo esto, sino que, además, entrenar los sistemas de TAN requiere una gran cantidad de tiempo, lo que puede significar días, semanas o incluso meses. Además, aun cuando ya se ha entrenado el sistema de TAN, la parte de decodificación de los modelos puede ser demasiado lenta si se lleva a cabo en ordenadores personales o portátiles (Britz et al., 2017: 01; Forcada, 2017: 295-300; Wu et al. 2016: 01). Por otro lado, es cierto que existen muchas herramientas de TAN con licencias que casi cualquier usuario podría costearse (o incluso gratuitas), como Nematus, AmuNMT o Neural Monkey. El problema yace en que suelen ser bastante complicadas de instalar y de configurar (Forcada, 2017: 301).

La TAN presenta también problemas de tipo lingüístico. Por un lado, la TAN trabaja bien con oraciones más cortas ya que los sistemas codifican la información contenida en las oraciones en LO en vectores numéricos. El problema surge cuando aparecen oraciones largas, sobre todo cuando estas son más largas que las oraciones del corpus de entrenamiento y si dichas oraciones contienen vocabulario que el sistema desconoce (Bahdanau, Cho y Bengio, 2014; Cho,

Merriënboer, Bahdanau y Bengio, 2014; Pouget-Abadie, Bahdanau, Merriënboer, Cho y Bengio, 2014). Según Bahdanau, Cho y Bengio (2014: 01):

A potential issue with this encoder–decoder approach is that a neural network needs to be able to compress all the necessary information of a source sentence into a fixed-length vector. This may make it difficult for the neural network to cope with long sentences, especially those that are longer than the sentences in the training corpus.

Es decir, que la calidad de los resultados de traducción se deteriora en el modelo codificador-decodificador si las oraciones de entrada son más extensas. La hipótesis de Cho, Merriënboer, Bahdanau y Bengio (2014) es que esto sucede porque los vectores no tienen capacidad suficiente para codificar oraciones largas con estructuras y significados complicados, por lo que, para codificar una secuencia de longitud variable, la red neuronal deberá «sacrificar» algunos componentes importantes de la oración de entrada para recordar otros.

Los sistemas de TAN suelen contar con un vocabulario limitado a unas 30.000 a 80.000 palabras, pero la traducción es un «open vocabulary problem» (Sennrich, Haddow y Birch, 2015), lo que significa, por ejemplo, que para idiomas con formación de palabras mediante aglutinación y composición (como es el caso del alemán), los sistemas de traducción deberán funcionar por debajo del nivel de palabra (Sennrich, Haddow y Birch, 2015). Es decir, que no se debería codificar palabras como *Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz* (en español «Ley para la transparencia de tareas de supervisión en el etiquetado de la carne vacuna») como si fuera un vector de longitud fija, sino que sería mejor dividirla en las diferentes palabras que la conforman de la siguiente manera «Rindfleisch | etikettierungs | überwachungs | aufgaben | übertragungs | gesetz».

La traducción neuronal también suele causar inconvenientes cuando debe gestionar grandes cantidades de vocabulario, ya que el entrenamiento y la decodificación se vuelven más complicados cuando la cantidad de palabras en LM es mayor, por este motivo el número de palabras en LM debe ser reducido. Este problema se suele intentar solucionar creando una *shortlist* (de entre 30.000 y 80.000 palabras) con las palabras más frecuentes de la lengua meta en cuestión (Bahdanau, Cho y Bengio, 2014; Jean, Cho, Memisevic y Bengio, 2014).

Aquellas palabras que no aparezcan en la lista se marcarán con un identificador de palabra desconocida «[UNK]». Esto funciona bien cuando en la oración meta hay pocas palabras desconocidas, pero, a medida que el número de palabras que el sistema no conoce aumenta en la oración de llegada, este enfoque se vuelve más problemático. Este problema se verá

agraviado en lenguas que son ricas en vocabulario y que tienen casos de declinación como es el caso del alemán (Jean, Cho, Memisevic y Bengio, 2014). Por último, uno de los inconvenientes más graves que presentan estos sistemas es que, en ocasiones, producen traducciones en LM donde no se han traducido todas las partes de la oración de entrada en LO, es decir comete omisiones, lo que puede dar lugar a traducciones erróneas (Wu et al., 2016: 01).

A partir de toda la información proporcionada en este apartado relativo a la traducción automática neuronal, donde se han expuesto sus orígenes, su funcionamiento o los modelos más innovadores (como el Traductor de Google), así como las principales ventajas e inconvenientes de estos sistemas neuronales, en el siguiente apartado se ha realizado una reflexión final sobre la TAN.

#### *2.3.2.8. Reflexiones finales sobre TAN*

La TAN se introdujo como un enfoque novedoso y prometedor con el potencial de solventar muchos de los problemas que presentaban los sistemas de traducción automática estadística. No obstante, la TAN actual sigue arrastrando algunos problemas en cuanto al rendimiento y la calidad de las traducciones, sobre todo si estas contienen palabras que desconoce, algo que limita el funcionamiento de estos sistemas neuronales para la traducción automática de textos en lenguas con pocos recursos o que sean ricas morfológicamente.

Por otro lado, Schulz, Aziz y Cohn (2018) nombran la importancia de tener en cuenta la variación, refiriéndose a las variaciones que surgen de las diferencias en el orden de las palabras en las lenguas ricas morfológicamente, como es el caso del alemán. Además, proponen como ejemplo el caso de una traducción de inglés al alemán. En inglés, las oraciones siguen un orden relativamente rígido, pero, tal vez, esa misma oración en alemán se pueda ordenar de diferentes maneras, sin embargo, en todas ellas se deberá respetar, por ejemplo, la relación entre sujeto y verbo, la posición de los casos del alemán, etc. Es decir, el sistema de TAN deberá aprender estas variaciones para traducirlas correctamente, por lo que se convierte en otro posible campo de estudio.

En secciones anteriores se han podido observar las principales ventajas e inconvenientes que presenta la TAN en comparación con la TAE. A continuación, se presentan algunas diferencias y similitudes adicionales que los sistemas neuronales tienen con respecto a los estadísticos. Entre las principales similitudes que encontramos en los dos sistemas se encuentra el hecho de que ambos se entrenan mediante corpus y que tanto en la TAN como en la TAE, la oración en LO se traduce a LM mediante el establecimiento de la probabilidad de que una sea la traducción de la otra (Forcada, 2017: 299).

Por otro lado, según Forcada (2017: 299), si se centra la atención en las diferencias entre la TAE y la TAN, encontramos que en TAN la probabilidad de que la oración en LM sea la traducción correcta, se calcula observando la probabilidad de cada palabra individual en LM, dada una oración original y las palabras en LM anteriores (el contexto). El proceso se lleva a cabo mediante redes neuronales, en que los pesos de las relaciones entre neuronas se entrenan de manera conjunta para conseguir mejores resultados.

En la TAE, se generan traducciones mediante el alineamiento de las traducciones de subsegmentos (normalmente, oraciones). Estos pares de subsegmentos, es decir, de oraciones en LO y oraciones en LM, se obtienen durante el entrenamiento utilizando corpus paralelos. En estos corpus se alinean las palabras en LO y sus traducciones en LM, mediante probabilidades estadísticas que se aprenden en el corpus. A continuación, se identifican las oraciones en LO y su correspondiente oración en LM, según las palabras que se hayan alineado. Así se podría afirmar que el entrenamiento en TAE ocurre en dos fases (alineación y extracción de oraciones) y no de forma conjunta (a través de una red neuronal) como en TAN (Bahdanau, Cho y Bengio, 2014; Forcada, 2017: 299).

Tras observar la evolución y el funcionamiento de los sistemas de traducción neuronal, es también imprescindible presentar algunas de las principales aplicaciones de los modelos de TAN, ya que éstas van más allá de la simple traducción automática de documentos de una lengua a otra. Además, la TAN tiene otras aplicaciones como la traducción asistida por ordenador (TAO), donde la TA se usa como cualquier otra herramienta dentro de los *softwares* de TAO. En estos entornos, la TA se utiliza como una herramienta para producir traducciones que después serán poseídas. Casacuberta y Peris-Abril (2017: 71) añaden además la traducción interactiva (TI) en que el traductor humano irá corrigiendo los errores que vaya detectando y que ha cometido la TA.

Estas correcciones se aprovecharán para mejorar la traducción en curso y las traducciones posteriores. Mencionan también que cada vez que se valida una traducción realizada mediante TA, se dispone de una nueva muestra para el aprendizaje, lo que permite mejorar los modelos de TA y evita que un mismo error se repita. Esto se lleva a cabo gracias a un proceso de adaptación por aprendizaje en línea (Casacuberta y Peris-Abril, 2017: 71). Otro uso muy común es el de la traducción mediante TA de reseñas en los portales de comercio electrónico, en que se traducen las reseñas y los comentarios para hacerlos más accesibles a los hablantes de diferentes lenguas (Castilho et al., 2017: 111). Además, en palabras de Castilho et al. (2017: 111):

A common use case in e-commerce consists in leveraging MT to make product descriptions, user reviews and comments (e.g. on dedicated forums) as widely accessible as possible, regardless of the customers' native language or country of origin.

Este último uso es en el que se centra la presente tesis, ya que una vez se traducen estas reseñas u opiniones es imprescindible poseerlas para que no haya errores que puedan dañar la percepción del producto o servicio.

### 2.3.3. La posesición

La TAN ha dado lugar a nuevos trabajos para los lingüistas como la posesición (PE). Según Allen (2003: 297), esta disciplina se ha utilizado en relación con diversas subespecialidades de la lingüística como el procesamiento de lenguaje natural, la TA, la corrección automática de errores, el reconocimiento óptico de caracteres, las memorias de traducción o los lenguajes controlados. Sin embargo, el término se ha utilizado sobre todo con relación a la TA y se ha definido como la corrección por parte de un lingüista o de un traductor humano de los textos traducidos mediante TA (Allen, 2003: 297; Veale y Way, 1997). Wagner la define como «post-editing entails correction of a pre-translated text rather than translation 'from scratch'» (Wagner, 1985: 01). Allen (2003: 297) añade que la tarea del poseedor consiste en editar, modificar y corregir texto previamente traducido por un sistema de TA, de una LO a una LM.

Allen (2003: 301) distingue entre varios niveles de PE que dependerán de la finalidad de los textos traducidos. Así, se pueden distinguir dos grados de PE en función de las exigencias del mercado, la parcial o *light post-editing* y la completa o *full post-editing* (Aranberri, 2014: 472; Candel-Mora y Borja-Tormo, 2017: 03). La posesición parcial consiste en realizar solamente aquellos cambios que sean imprescindibles para que el texto se pueda entender (cuestiones de terminología o gramaticales que dificulten la lectura). Por otro lado, la posesición completa consiste en eliminar todos y cada uno de los errores que pueda haber cometido el sistema de TA para conseguir una traducción final al nivel de la traducción humana (no solo cuestiones de terminología y gramática, sino también estilísticas y de fluidez). Será el cliente el que decida el nivel al que se tiene que llevar a cabo, para alcanzar el nivel de calidad determinado que necesita en su texto.

Candel-Mora y Borja-Tormo (2017: 02) aseguran que la posesición se ha estudiado desde diferentes perspectivas. Entre estas, distinguen los puntos de vista de la calidad, de las pautas de evaluación, del esfuerzo cognitivo, de la aceptabilidad de la producción de TA, de la combinación de estrategias como la pre-edición y el del uso de lenguajes controlados para

mejorar la traducibilidad. Adicionalmente, también se ha estudiado desde el punto de vista comercial para comprobar la productividad con ayuda de herramientas TAO. No obstante, todavía falta estudiar en profundidad aquellos enfoques que puedan predecir el esfuerzo necesario para llevar a cabo las tareas de PE (Candel-Mora y Borja-Tormo, 2017: 02).

Aunque esta profesión está cada vez más presente en el mercado de la traducción, hoy en día sigue habiendo muchas voces críticas, ya que, en ocasiones, esta se ha utilizado para recortar costes de traducción, lo que ha resultado en unas tarifas mucho más reducidas para los traductores profesionales (Sánchez-Gijón, 2016) e incluso en la supresión de su trabajo, ya que muchas empresas optan erróneamente por traducir sus contenidos con TA sin posterior revisión por un profesional.

Por el contrario, en un estudio llevado a cabo por Vasconcellos (1988: 145) con un total de 15 poseedores que trabajaban con los sistemas ENGSPAN y SPANAM, de la Organización Panamericana de la Salud, se llegó a diversas conclusiones en cuanto a la posesición de documentos. En primer lugar, los poseedores aseguraron que las competencias necesarias en esta disciplina se desarrollan gradualmente y que, a medida que poseitan textos, los prejuicios que se pueden tener sobre la posesición van desapareciendo. Además, los lingüistas se sienten más seguros cuando ya han poseitado aproximadamente 100.000 palabras (equivalente a un mes de trabajo). Estos poseedores aseguraron también que la posesición puede llegar a ser más relajante y menos fatigante que traducir un texto desde cero. Aranberri (2014: 474) y Toledo-Báez (2018: 25) se muestran de acuerdo con la afirmación de que la PE ayuda a los lingüistas a ser más productivos.

También se ha afirmado que la TAN ha ayudado a que el proceso de posesición sea más eficiente, ya que se han obtenido mejores puntuaciones que con la TAE según métricas como BLEU. Según estas métricas, las traducciones generadas mediante TAN producen textos con mayor fluidez y naturalidad, por lo que, en teoría deberían requerir menor tiempo de posesición. No obstante, según estudios recientes como el de Sánchez-Gijón, Moorkens y Way (2019: 32), la mayor parte de los traductores no percibe ninguna mejora en cuanto a su productividad gracias a la TAN. En dicho estudio, los traductores afirmaron que las traducciones con TAN consiguen buenos resultados en cuanto a la distancia de edición y a las correcciones requeridas, algo que, en general, se traduce en un menor esfuerzo de posesición. No obstante, el tiempo que se invierte en la posesición de traducciones generadas con TAN es mayor.

Algo en lo que parece haber consenso es que la posesición debe ser llevada a cabo por traductores experimentados (McElheney y Vasconcellos, 1988: 142; Zaretskaya, 2017: 116). Esto se puede observar en la afirmación de McElheney y Vasconcellos (1988: 142):

[...] postediting is and will continue to be for a long time the work of the profesional translator. There are strong arguments in favor of using experienced translators for this task. The translator is the one best able to pick up errors in the machine translation (e.g., misparsed or unparsable ambiguities), he has a fund of knowledge about the cross-language transfer of concepts, and he has technical resources at his disposal which he knows how to use in the event of doubts. Moreover, for the very reason that translators are best suited to the task, the more experienced they are, the more effective they will be.

No obstante, esta es una disciplina relativamente nueva y muchos lingüistas todavía no saben muy bien cómo llevarla a cabo. Por este motivo, es necesario que exista formación en posesición tanto en las universidades como por parte de las empresas (Zaretskaya, 2017: 118).

#### 2.3.4. Métricas de calidad de TAN

Evaluar los sistemas de TAN es esencial para que estos se puedan mejorar, sin embargo, todavía no hay una definición consensuada de calidad de TAN entre la academia y la industria de la traducción y la localización (Castilho et al., 2017: 109). No obstante, en los últimos años, se ha realizado un mayor esfuerzo en unificar el proceso de evaluación gracias a los estándares de calidad ISO/TS 11669:2012 y ISO 18587:2017 y a algunas asociaciones como la desaparecida LISA o como TAUS (Van der Meer et al., 2017). Según Lommel, Uszkoreit y Burchardt (2014: 455) uno de los principales problemas de estas métricas de calidad es la falta de estandarización, la cual Doherty (2017: 132) atribuye al hecho de que algunas de las métricas de calidad de traducción automática se basan en el texto original y otras en el texto meta. Además, los diferentes tipos de texto y pares de lenguas, así como la complejidad del proceso de traducción, dificultan la definición de un objetivo de calidad unificado (Doherty, 2017: 132).

El nivel de calidad de la TAN se puede evaluar a través de mecanismos automáticos de evaluación, como BLEU, o con métricas manuales como MQM (Multidimensional Quality Metrics), desarrollada gracias al proyecto QTLaunchPad, financiado por la Unión Europea, o el Dynamic Quality Framework (DQF) de TAUS, que es el modelo más utilizado en la actualidad (Lommel, Uszkoreit y Burchardt, 2014: 461).



Según Valli (2015: 127), el modelo dinámico de TAUS (Van der Meer et al., 2017) gira en torno a tres parámetros de evaluación (*utility, time* y *sentiment*) y, cada uno de ellos contará con mayor o menor relevancia según el contenido que se vaya a traducir. TAUS define «utilidad» (*utility*) como la funcionalidad del contenido, «tiempo» (*time*) como la velocidad a la que se llevó a cabo la traducción y «sentimiento» (*sentiment*) como la forma en que los usuarios perciben la imagen de cierta marca (O'Brien et al., 2011). Este modelo evalúa algunas tipologías de error como la precisión (*accuracy*), la fluidez (*fluency*), la terminología, el estilo, el diseño, las convenciones propias de los *locales* o la veracidad (*verity*) (Van der Meer et al., 2017). Además, el modelo DQF cuenta con varios niveles de gravedad para las diferentes categorías de error.

Es evidente que la TA ha evolucionado desde sus inicios hasta la actualidad y que ha creado nuevos campos de estudio y de trabajo no solo para ramas técnicas como la informática, sino también para la traducción. Sin embargo, hasta llegar al punto en que está hoy en día, ha pasado bastante tiempo. En la siguiente sección, se tratarán temas como el origen de la TA; sus primeras formas, características y usos hasta el día de hoy, entre otras cuestiones. Así, se podrá observar cómo y cuándo han aparecido los diferentes sistemas de TA hasta llegar a los modelos actuales. Del mismo modo, a través de esta revisión histórica de los orígenes de la TA, que comenzará en los años 30 y finalizará en la actualidad, se presentarán los principales precursores y las diferentes perspectivas desde las que se ha estudiado este campo de la traducción.

### 2.3.5. Orígenes y evolución de la TA

En el presente apartado, ha realizado un recorrido histórico a través de los orígenes de la TA, desde su aparición aproximadamente en los años 30 hasta el presente. Como se ha comentado anteriormente, la TA no nace como un sistema sofisticado de traducción neuronal, ni siquiera como un sistema de TA estadística, sino que esta nace incluso antes que los primeros ordenadores. De esta manera, se han estudiado los principales sistemas de TA y su funcionamiento y se ha elaborado una recopilación de los más utilizados, de manera que se pueda observar su evolución a lo largo del tiempo hasta la actualidad.

#### 2.3.5.1. Años 30: Artsrouni y Trojanskij, precursores

Los primeros en acuñar el término «máquinas de traducir» o *translating machines* fueron el francés Georges Artsrouni y el ruso Petr Trojanskij, los cuales solicitaron patentes a mitad de los años 30 (Hutchins, 2001; Oliver, 2014: 03; Tertoolen, 2012: 144). La de Trojanskij propuso un diccionario automático bilingüe, además de un esquema para codificar e interpretar funciones gramaticales interlingüísticas mediante el uso de símbolos universales basados en el esperanto.

La de Artsrouni era más bien una máquina de uso general que funcionaba también como un diccionario automático multilingüe (Hutchins, 2001: 09).

La máquina de Artstrouni, patentada bajo el nombre de «cerebro mecánico» (Hutchins, 2004: 13), tenía múltiples aplicaciones, entre las cuales destacó la de traducción desde un idioma a tres idiomas distintos. La máquina estaba formada por un teclado que activaba el lector, un mecanismo de búsqueda o seleccionador, un mecanismo de producción y un dispositivo de memoria que almacenaba palabras en cuatro lenguas (Henisz-Dostert, Macdonald y Zarechnak, 2011: 07; Hutchins, 2004: 14) (Ilustración 14).

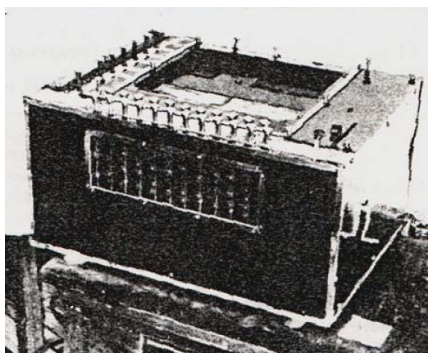


Ilustración 14. «Cerebro mecánico» de Artsrouni. Imagen extraída de Hutchins, 2004: 13.

Artsrouni no era lingüista y no conocía algunos de los problemas más comunes de la traducción como la polisemia, las frases hechas o la ambigüedad sintáctica, por lo que no creó esta máquina como un mecanismo de traducción automática, ni siquiera para obtener traducciones de gran calidad, sino para obtener traducciones rápidas y aproximadas (Hutchins, 2004: 15).

Por otra parte, Trojanskij quería crear una «máquina de traducir» o «una máquina para seleccionar y escribir palabras cuando se traduce de una lengua a otra o a varias lenguas simultáneamente» (Hutchins, 2004: 17). La máquina consistía en una mesa ligeramente inclinada, sobre la cual se movía una cinta con perforaciones (Ilustración 15):

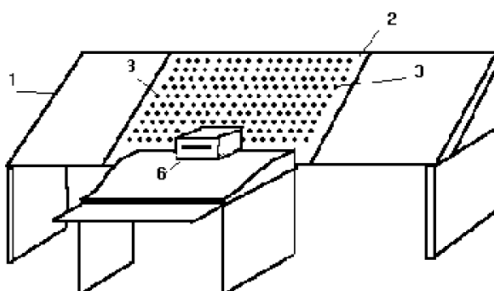


Ilustración 15. «Máquina de traducir» de Trojanskij. Imagen extraída de Hutchins, 2004: 18.

La cinta con perforaciones (número 3 en la Ilustración 15) era un diccionario con entradas en seis lenguas en columnas en paralelo (Henisz-Dostert, Macdonald y Zarechnak, 2011: 07). Según Hutchins (1995: 04), Trojanskij visualizó una traducción automática a tres niveles:

[...] first, an editor knowing only the source language was to undertake the 'logical' analysis of words into their base forms and syntactic functions; secondly, the machine was to transform sequences of base forms and functions into equivalent sequences in the target language; finally, another editor knowing only the target language was to convert this output into the normal forms of his own language.

Para utilizar la máquina, la persona que la utilizara debía localizar una palabra en LO y ajustar la cinta móvil para mostrar la palabra en las diferentes LM. Una vez encontradas las equivalencias de las traducciones, un revisor debía editar el texto para extraer el significado de la traducción y eliminar cualquier incorrección (Hutchins, 2004: 18). De esta manera, se podría decir que este es uno de los primeros pasos hacia la posesición de textos, ya que un lingüista debía revisar los resultados obtenidos mediante la TA de la máquina de Trojanskij.

Aunque ambos patentaron sus máquinas de traducir en el 1933, tanto la máquina de Artsrouni como la de Trojanskij fueron desconocidas hasta finales de los años 50 (Hutchins, 2004: 16), momento en que el ordenador ya había nacido (Hutchins, 2014: 01) y década marcada por la influencia del gran experimento de Turing y la criptografía. A continuación, se ha llevado a cabo un recorrido histórico de la década que abarca desde los años 40 hasta los 50, y se ha investigado a algunos de los grandes referentes del momento, como Weaver y la Rockefeller Foundation, Booth o Bar-Hillel.

#### 2.3.5.2. *Años 40 - 50: de la criptografía al auge de la investigación en TA*

Aunque la idea de utilizar ordenadores para traducir lenguajes naturales surge en los años 40 (Hutchins, 2001: 05), las teorías de que las máquinas podrían tener inteligencia y comunicarse se remontan a Alan Turing, matemático británico conocido, entre otros aspectos, por descifrar «Enigma», la máquina para encriptar mensajes de radio de los nacionalsocialistas alemanes durante la II Guerra Mundial (Enciclopedia Britannica, 2023). Este tuvo gran influencia en el desarrollo de los ordenadores y en la teoría de que estos pudieran ser máquinas inteligentes, por lo que se le ha considerado un pionero de la Inteligencia Artificial (Hutchins, 1986).

Para probar si las máquinas disponían de inteligencia, Turing creó el llamado *Imitation Game* o «Juego de Imitación», diseñado para responder a la pregunta de si las máquinas podían pensar. El experimento consistía en lo siguiente (Turing, 1950: 433):

It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A."

El papel del hombre (A) era hacer que el interlocutor (C) se equivocara en su toma de decisión. Como se ha mencionado anteriormente, Turing quería responder a la pregunta «¿pueden pensar las máquinas?», por este motivo añade un tercer participante al juego, la máquina, la cual debía hacer el papel de hombre (A) y engañar al interlocutor (C). El objetivo era responder a la pregunta «¿qué pasaría si una máquina entra en el juego con el rol de A?». Es decir, quiere comprobar si el interlocutor (C) se equivocará al tomar decisiones con lo que le diga la máquina, haciéndose pasar por A, tanto como se equivocaba cuando jugaba contra el hombre (A) (Turing, 1950: 433).

Fueron muchos los que siguieron los pasos de Turing en el estudio de la criptografía, las matemáticas y la informática. Entre ellos Warren Weaver, de la Rockefeller Foundation, y Andrew D. Booth, un ingeniero británico, los cuales proponen por primera vez la posibilidad de que los ordenadores se puedan utilizar para traducir lenguajes naturales (Hutchins, 1995: 434; Hutchins, 2001: 09). Por un lado, Booth quiso explorar la posible mecanización de un diccionario bilingüe con la colaboración de Richard H. Richens, el cual había estado experimentando para conseguir traducciones automáticas palabra por palabra rudimentarias para resúmenes científicos (Hutchins, 1995: 434).

No obstante, según Hutchins (1995, 435), es Weaver en su memorándum de 1949 (publicado en 1955) el que saca a la luz la idea de traducción automática, esboza diferentes posibilidades sobre esta y propone distintos métodos como técnicas criptográficas (ya utilizadas en la II Guerra Mundial), métodos estadísticos, la teoría de la información de Shannon y la exploración de la lógica y las características del lenguaje.

Weaver menciona también la problemática de la traducción en el contexto de la comunicación entre personas hablantes de distintas lenguas y reconoce las dificultades semánticas que plantea (Hutchins, 2005: 01; Weaver, 1955: 04). Por lo que sugiere la posibilidad de utilizar máquinas para traducir texto, aunque fuera solo material científico, ya que en este las dificultades semánticas son menores. Según Weaver (1955: 04), no era imprescindible que la traducción fuera perfecta, solo era necesario que los resultados fueran inteligibles.

Finalmente, en su memorándum, Weaver (1955: 04) planteó la posibilidad de que la traducción se pueda tratar como un problema de criptografía. Esta teoría se inspira en Claude Shannon (padre de la Teoría Matemática de la Comunicación), que también trabajó en el campo de la criptografía y planteó junto con Turing la posibilidad de una máquina que simulara el cerebro humano (Guizzo, 1999: 20). Wiener no se mostró tan optimista sobre las posibilidades de la traducción automática, ya que el lenguaje puede llegar a ser muy emocional y vago, e incluso contar con connotaciones demasiado complejas como para que un ordenador las pudiera traducir correctamente (Weaver, 1955: 05). En palabras de Weaver (1955: 05):

Second - as to the problem of mechanical translation, I frankly am afraid the boundaries of words in different languages are too vague and the emotional and international connotations are too extensive to make any quasi mechanical translation scheme very hopeful [...]

No obstante, Andrew D. Booth, profesor en la Universidad de Londres, defendió la posibilidad de que un ordenador pudiera traducir de una lengua a otra. Según Weaver (1955: 06), Booth expresa lo siguiente:

A concluding example, of possible application of the electronic computer, is that of translating from one language into another. We have considered this problem in some detail and it transpires that a machine of the type envisaged could perform this function without any modification in its design.

En resumen, el memorándum de Weaver aseguraba que los idiomas son solo códigos, cuyo significado se podía descifrar si estos se descriptaban y analizaban con un ordenador, al igual que hizo Turing durante la guerra. No obstante, aunque el memorándum de Weaver fue el documento que puso de manifiesto el ímpetu de algunos investigadores en estudiar la posibilidad de que los ordenadores pudieran traducir lenguajes naturales, la primera persona en dedicarse a tiempo completo al estudio de este campo en 1951, centrar su trabajo en estudiar las posibilidades de la TA y planificar futuras investigaciones en esta área fue Yeshoshua Bar-Hillel (Hutchins, 1986; Hutchins, 2001: 09).

En 1951, Bar-Hillel redactó un informe sobre el estado de la TA en aquel momento en que afirmaba lo siguiente (Bar-Hillel, 1951: 229):

It seems obvious that fully automatic MT, i.e. one without human intervention between putting the foreign text into the reading organ of the mechanical translator and reading off its output, is achievable only at the price of inaccuracy, if only for the

reason that no method is feasible, for the time being, by which the machine would eliminate semantical ambiguities.

Lo que Bar-Hillel planteó en su informe era que, en aquel momento, la TA no era capaz de traducir textos de manera precisa sin ayuda de un humano, puesto que no podía eliminar las ambigüedades propias de las distintas lenguas. Además, también afirmó que no sería posible alcanzar la llamada *Fully Automatic High-Quality Translation* (FAHQT o TA totalmente automática de alta calidad) con la tecnología de aquel momento. En dicho informe, Bar-Hillel defiende la importancia de implementar una estrategia de posesición para eliminar dichas ambigüedades (Bar-Hillel, 1951: 230; Hutchins, 2001: 09). Bar-Hillel (1951: 230) expresa:

It appears that a post-editor is indispensable for elimination of semantical ambiguities [...] He should be able to produce out of the raw output of this hypothetical machine a readable translation in a fraction of the time it would take a bilingual expert to produce a translation with the conventional procedure.

En junio de 1952, se celebró la primera conferencia sobre esta temática y que estuvo patrocinada por la Rockefeller Foundation (Gordin, 2015: 261; Hutchins 1986; Hutchins, 2001: 09). En esta, se confirmó que la traducción completamente automática todavía no era posible sin la ayuda de un humano en el paso previo o posterior a la TA, es decir, sin pre o posesición (Hutchins, 2001: 09). No obstante Bar-Hillel, vio potencial en el uso de la TA en situaciones donde se pudiera utilizar vocabulario más restringido, por ejemplo, para textos dirigidos a pilotos o a meteorólogos; con idiomas como el inglés simplificado o con lenguajes auxiliares como el esperanto o la interlingua (Hutchins, 1986).

Uno de los resultados de esta conferencia fue la creación de un grupo de investigación en TA en la Universidad de Georgetown, que empezó a trabajar en la primera demostración pública de un sistema de TA, que salió a la luz en el año 1954 (Dostert, 1957: 04; Gordin, 2015: 261; Hennisz-Dostert, Macdonald y Zarechnak, 2011: 07; Hutchins, 1986; Hutchins, 2001: 10; Hutchins, 2005: 01).

Este sistema traducía un corpus de 49 oraciones del ruso al inglés utilizando un vocabulario compuesto por 250 palabras y solo 6 reglas gramaticales y, aunque tuvo poco valor científico, consiguió despertar el interés en este tipo de experimentos, los cuales se empezaron a financiar a gran escala en los EE. UU., así como también en la antigua Unión Soviética. Por razones políticas y militares la mayor parte de la investigación en los EE. UU. se basó en sistemas de TA de ruso a inglés y en la antigua Unión Soviética fue en TA de inglés a ruso (Gordin, 2015: 261; Hennisz-Dostert, Macdonald y Zarechnak, 2011: 07; Hutchins, 2001: 11).

En este momento, también se publicaron por primera vez la revista especializada «Machine Translation» (creada en 1953 por William Locke y Victor Yngve, sucesores de Bar-Hillel en el MIT), además de tesis doctorales o libros, entre otros (Hutchins, 1986; Hutchins, 2001: 10). Uno de estos libros fue *Machine Translation of Languages: Fourteen Essays*, una colección de artículos editados por Locke y Booth que se convirtió en el primer libro publicado dedicado en su totalidad a la TA. Este libro está compuesto por catorce artículos estructurados de la siguiente forma (Locke y Booth, 1955):

1. *Translation*: prólogo escrito por Warren Weaver del que se hablaba anteriormente.
2. *Some Methods of Mechanized Translation*: escrito por R. H. Richens y A. Donald Booth. Sobre los principios y métodos de la que llamaban «traducción mecanizada».
3. *The Design of an Automatic Russian-English Technical Dictionary*: escrito por Anthony G. Oettinger. En este, el autor habla del diseño de un diccionario técnico del ruso al inglés para su implementación en un sistema de TA.
4. *A Preliminary Study of Russian*: escrito por Kenneth E. Harper. Muestra los resultados de un estudio preliminar sobre TA de ruso.
5. *Some Problems of the "Word"*: escrito por William E. Bull, Charles Africa y Daniel Teichroew. Trata sobre la problemática del concepto «palabra» en TA.
6. *Speech Input*: escrito por W. N. Locke. Trata sobre la traducción de la lengua hablada, en este caso el inglés.
7. *Storage Devices*: escrito por A. Donald Booth. Sobre las unidades de almacenamiento de las que depende la TA.
8. *The Georgetown-I.B.M. Experiment*: escrito por Leon E. Dostert. El autor realiza una revisión del origen del experimento de TA realizado en 1954 y de la metodología utilizada para llevarlo a cabo.
9. *The Mechanical Determination of Meaning*: escrito por Erwin Reifler. Trata sobre las dificultades que plantea la automatización de la traducción.
10. *Model English*: escrito por Stuart C. Dodd. Trata sobre las reglas que permiten a los sistemas de TA traducir desde una lengua modelo (en este caso el inglés) o hacia una lengua modelo.
11. *A Practical Development Problem*: escrito por James W. Perry. Trata sobre los costes y la inversión de capital que se debe realizar si se quiere traducir con sistemas de TA.
12. *Idioms*: escrito por Yehoshua Bar-Hillel. Sobre las dificultades de la TA al traducir este tipo de oraciones.
13. *Some Logical Concepts for Syntax*: escrito por Luitgard y Alex Wundheiler.

14. *Syntax and the Problem of Multiple Meaning*: escrito por Victor H. Yngve. Tanto este artículo como el anterior hacen referencia a los problemas sintácticos que enfrenta la TA.

De esta forma, se comenzaron a publicar más estudios en y, en octubre de 1956, el MIT celebró la primera conferencia internacional sobre TA con asistentes de EE. UU., Canadá y Gran Bretaña y con artículos recibidos de la Unión Soviética. A partir de este mismo año, se fundaron grupos de investigación en TA en todo el mundo y en EE. UU., el gobierno, el ejército y otras fuentes privadas crearon becas para fundar dichas investigaciones (Hutchins, 1986). Según Hutchins (2001: 11-12), en 1958, apareció el primer proyecto práctico, dirigido por Gilbert King en IBM Corporation, y se creó un sistema que realizaba traducciones rudimentarias para su uso en las fuerzas aéreas estadounidenses, el cual en 1970 se sustituyó por Systran.

No obstante, y como se describe a continuación, en 1964, la opción de sistemas de TA que fueran realmente utilizables y operativos parecía todavía imposible, por lo que los financiadores y patrocinadores de la investigación en TA crearon un comité que acordó reducir la financiación de los proyectos de investigación en TA (Informe ALPAC, 1966: 32, Hutchins, 1986; Hutchins, 2005: 02; Oliver, 2014: 03; Slocum, 1985: 01; Tertoolen, 2012: 145).

#### 2.3.5.3. *Años 60-70: el Informe ALPAC y sus consecuencias*

En los años 60, durante Guerra Fría entre EE. UU. y la URSS, la actividad de estos dos países en cuanto a TA se había centrado en la traducción de inglés al ruso y viceversa de documentos científicos y técnicos para un número reducido de usuarios, que pudieran utilizarlos, aunque el resultado de la TA fueran traducciones rudimentarias, cuyo principal objetivo fuera el acceso rápido a la información (Hutchins, 2005: 02). Además, Bar-Hillel (1960: 93) criticó la creencia generalizada de que el objetivo de la investigación en TA debería ser la de crear sistemas TA de alta calidad cuyos resultados fueran idénticos a los de un traductor humano.

Aunque en los años 50 la TA había suscitado gran interés y había recibido un gran apoyo, esto cambió a partir de la década de los 60 (Slocum, 1985: 01). En 1964, los principales patrocinadores estadounidenses de las investigaciones en TA, preocupados por el lento avance de estos estudios, formaron el *Automatic Language Processing Advisory Committee* (ALPAC por sus siglas en inglés), del cual surgió el conocido «Informe ALPAC» que concluyó que la TA era en aquel momento «slower, less accurate and twice as expensive as human translation» (Hutchins, 2005: 02) y que «there is no immediate or predictable prospect of useful machine translation» (Informe ALPAC, 1966: 32). Es decir, que la TA no se consideraba una buena opción porque no



alcanzaba la calidad de las traducciones llevadas a cabo por los humanos, y además era mucho más costosa.

Así, el informe ALPAC puso fin a la investigación en TA tanto en EE. UU., como en la Unión Soviética y en Europa durante al menos una década. De esta manera, se concluyó que ya no se necesitaría más inversión en TA, por lo que se cancelaron muchos proyectos sobre TA en EE. UU. Por eso, los esfuerzos se comenzaron a centrar en la creación de herramientas para traductores, como diccionarios, herramientas de traducción asistida por ordenador (TAO) y avances básicos en lingüística computacional (Hutchins, 2001: 06; Hutchins, 2005: 02; Hutchins, 2015: 04; Oliver, 2014: 03; Slocum, 1985: 01; Tertoolen, 2012: 145).

Al principio de la década de los 70, se creó Systran, un sistema de TA, utilizado por las fuerzas aéreas estadounidenses, para traducir documentos militares científicos y técnicos del ruso al inglés y, en 1974, este se instaló en la NASA (Gomes de Oliveira y Anastasiou, 2011: 122; Hutchins, 2001: 12; Hutchins, 2005: 02; Hutchins, 2015: 04; Oliver, 2014: 04; Tucker, 1984: 322).

En 1976, la Comisión Europea compró una versión de Systran para traducir textos de inglés a francés y, más tarde, también para la mayoría de las lenguas de la Unión Europea, ya que se tenían que traducir grandes cantidades de texto (Gotti, Langlais y Lapalm, 2014: 400; Hutchins, 2001: 17; Hutchins, 2005: 02; Langlais et al, 2005: 84; Oliver, 2014: 04). De manera paralela, un grupo de investigación de la Universidad de Montreal (Canadá), llamado TAUM, desarrolló el sistema de TA llamado TAUM-MÉTÉO, que se utilizó para traducir partes meteorológicos primero de inglés a francés y, más tarde, de francés a inglés (Gotti, Langlais y Lapalm, 2014: 400; Langlais et al, 2005: 84).

A principios de los años 70, se siguieron tres líneas principales de investigación en TA. Por un lado, se crearon las primeras herramientas de traducción asistida por ordenador, además se desarrollaron sistemas de TA con ayuda de traductores humanos y, por último, se llevaron a cabo investigaciones teóricas para mejorar los métodos y los sistemas de TA (Hutchins, 2001: 07). A mediados de los 70, comenzó a crecer la demanda de sistemas de TA por parte de diferentes organizaciones. En Europa, por ejemplo, se comenzaron a necesitar más traducciones a causa del creciente comercio internacional, por lo que, en este caso, la demanda se centró en sistemas de traducción asistida por ordenador rentables que ayudaran a traducir la documentación técnica y comercial a los principales idiomas del comercio internacional (Hutchins, 2005: 02).

#### 2.3.5.4. Años 80-90: comercialización de los sistemas de TA y de herramientas TAO

Dada la creciente necesidad de traducción en todo el mundo, en los años 80, se comenzaron a comercializar una gran variedad de sistemas de TA en distintos países. Además, los ordenadores personales empezaron a ser cada vez más frecuentes, por lo que los usuarios también demandaron sistemas que ayudaran con la gestión lingüística de sus documentos y que además fueran económicos (Craciunescu, Gerding-Salas, Stringer-O'Keeffe, 2004; Hutchins, 2005: 03), no solo sistemas de TA, sino también herramientas TAO.

A principios de los años 80, aparecieron los primeros sistemas de TA para ordenadores personales que se utilizarían para realizar traducciones automáticas rudimentarias con propósitos informativos (sin necesidad de posesición) o para producir traducciones-borrador para textos con otros propósitos que sí que necesitaran posesición. En palabras de Hutchins (2001: 04):

When the first versions of MT systems appeared for personal computers (in the early 1980s) it was widely assumed that they would be used in much the same way as the mainframe systems had been, i.e. either to produce 'rough' (unrevised) versions for information purposes, or 'draft translations' for later revision and refinement. In both cases, it was also widely assumed that the principal recipients of MT output would be translators or at least people with good knowledge of both source and target languages; and, in the case of large organisations, it was expected that most would be professionally trained translators.

Pensando en esto, algunas organizaciones asumieron que los principales usuarios de este tipo de sistemas de TA serían traductores profesionales (Hutchins, 2001: 04). Este aumento de la demanda tanto de herramientas TAO como de sistemas de TA para ordenadores personales se produjo por 4 razones principales (Hutchins, 2001: 04):

1. La aparición y comercialización de estaciones de trabajo o *workstations* diseñadas específicamente para traductores profesionales.
2. Los sistemas basados en ordenador empezaron a ser utilizados por usuarios sin interés en la traducción como tal. Es decir, los utilizaban como mecanismos de asistencia a la comunicación, por lo que la calidad de la traducción era secundaria.
3. La expansión de las redes de telecomunicaciones y de Internet. Se hizo posible la comunicación inmediata en diferentes idiomas, lo que evidenció la necesidad de sistemas y servicios de traducción que pudieran dar soluciones en tiempo real a los problemas lingüísticos planteados por el creciente número de textos electrónicos de toda clase.

4. La creciente aparición de recursos de información y bases de datos multilingües creó la necesidad de búsquedas en diferentes idiomas online y dispositivos de acceso con módulos de traducción integrados.

En este periodo también surgieron algunas de las principales empresas en el panorama de la TA comercial para suplir estas necesidades. Algunas de ellas fueron ALPS, Weidner, Linguistic Products o Globalink, en Norteamérica y en Europa, y Sharp, NEC, Oki, Mitsubishi y Sanyo en Japón. Aunque también aparecieron otros sistemas en China, Taiwán, Corea, Europa del Este o en la Unión Soviética (Craciunescu, Gerding-Salas, Stringer-O'Keeffe, 2004; Hutchins, 2005: 03; Tertoolen, 2012: 146).

Los primeros sistemas en aparecer fueron los sistemas americanos Weidner (1981) y Alps (1983). El sistema ALPS proporcionaba ayuda en los siguientes tres niveles: procesamiento de texto multilingüe, diccionarios automáticos y consultas terminológicas, y traducción interactiva (traducciones automáticas rudimentarias y tipo borrador para revisión por parte de los traductores). El sistema Weidner estaba disponible en gran cantidad de pares de lenguas, pero fue particularmente popular para traducciones de japonés a inglés (Hutchins, 2001: 18-19). De esta manera, se aprecia cómo, aparte de Systran, que a partir de los 80 opera en más idiomas, aparecen otros sistemas de TA. Por ejemplo, tres de los más conocidos fueron Logos, Metal y los sistemas creados de manera interna por la Organización Panamericana de la Salud (OPS), llamados SPANAM y ENGSPAN (Hutchins, 2005: 03; Hutchins, 2015: 09; Slocum, 1985).

Logos, se empezó a desarrollar en los 70 por la empresa Logos Corporation en un proyecto dirigido por Bernard E. Scott, que consistía en la creación de un sistema de TA que tradujese manuales de aviación de inglés a vietnamita. Más adelante, el proyecto se amplió para introducir también traducción de alemán a inglés y Logos se posicionó como el principal competidor de Systran a partir de 1982. Más tarde, se incluyeron otros pares de lenguas como inglés a francés, alemán, italiano y español, o de alemán a francés o italiano. Aunque en un principio Logos se diseñó como un sistema basado en traducción directa, más tarde se incluyeron sistemas para el registro y la aplicación de reglas semánticas (Craciunescu, Gerding-Salas, Stringer-O'Keeffe, 2004; Hutchins, 1995: 11; Hutchins, 2001: 17; Hutchins, 2005: 03; Slocum, 1985).

El sistema Metal apareció a finales de los años 80 como un sistema de TA comercial que traducía de alemán a inglés. Este se desarrolló en la Universidad de Texas y, más tarde, se comercializó también para otras lenguas como el holandés, el francés y el español. Durante la mayor parte de los años 80, Siemens utilizó este sistema de manera interna y, al final de la década, el sistema Metal se empezó a comercializar con poco éxito. Otros ejemplos de compañías europeas que

utilizaban este sistema fueron Boehringer Ingelheim, SAP, Philips, y el Union Bank of Switzerland (UBS) (Hutchins, 2001: 18; Hutchins, 2003: 10). Tanto Systran como Logos y Metal fueron sistemas diseñados en principio para ámbitos de uso generales. No obstante, los diccionarios de estos sistemas se acabaron utilizando para ámbitos de trabajo más específico (Hutchins, 2001: 18).

De manera similar, aparecieron los sistemas desarrollados internamente por la Organización Panamericana de la Salud (OPS), los cuales traducían de español a inglés (llamado Spanam) y de inglés a español (llamado Engspan). Estos fueron desarrollados por Muriel Vasconcellos y Marjorie Leon y eran sistemas de traducción general que no se ceñían a un vocabulario específico o a un tipo de texto en concreto y cuya principal fortaleza fueron los diccionarios del campo de las ciencias sociales aplicadas al ámbito de la salud (Hutchins, 2001: 18; Hutchins, 2003: 12; Hutchins, 2015: 09; Tucker, 1984: 322).

Según Hutchins (2005: 03), de manera paralela, en los años 80, algunas compañías informáticas japonesas como Fujitsu, Hitachi, NEC, Sharp o Toshiba empezaron a desarrollar sistemas de traducción (tanto automática como TAO) de inglés a japonés y viceversa. Estos sistemas, que solían estar restringidos a campos temáticos específicos, sobre todo a la informática o a las tecnologías de la información, no ofrecían traducciones totalmente funcionales, por lo que necesitaban de asistencia de un traductor profesional tanto a nivel de pre-edición (etapa preparatoria del texto) como a nivel de posesición (etapa de revisión) (Hutchins, 2001: 18).

Hutchins (2015: 07; 2001: 19) asegura que el proyecto más reconocido de los años 80 fue el proyecto «Eurotra», un sistema de TA de segunda generación elaborado por la Unión Europea y en el cual estuvieron involucrados equipos de cada país miembro (Hutchins, 2005: 03; Hutchins, 2015: 06; van Slype, 1979: 11). Eurotra se desarrolló con el propósito de crear un sistema de transferencia multilingüe para todos los idiomas de los estados miembros de la Unión y se hizo basándose en el supuesto de que el enfoque de traducciones directas que ofrecía Systran era limitado (Hutchins, 2001: 19-20; Hutchins, 2015: 07; Tucker, 1984: 323).

En un principio, se esperaba que Eurotra terminara sustituyendo a otros sistemas como Systran que la Comisión había adquirido y que seguía desarrollando de manera interna. Además, durante su desarrollo, el proyecto estimuló la investigación en lingüística computacional y lingüística teórica, especialmente en Países Bajos, Bélgica, Dinamarca, Alemania y Gran Bretaña, donde se avanzó mucho en el establecimiento de las bases teóricas de la TA. Sin embargo, pese al trabajo y al tiempo invertidos en Eurotra, este proyecto no ofreció los resultados que los

patrocinadores esperaban y, en 1992 se paralizó definitivamente (Hutchins, 1995: 444; Hutchins, 2003: 08).

A finales de los años 80, se volvió a despertar el interés por los sistemas basados en la conocida como «interlingua», en parte, por la cada vez más frecuente investigación en inteligencia artificial (Hutchins, 1995: 445; Hutchins, 2015: 07). Según Hutchins (1995: 433), «Interlinguas may be based on an artificial language, an auxiliary language such as Esperanto, a set of semantic primitives presumed common to many or all languages, or a 'universal' language-independent vocabulary». Es decir, una lengua intermedia que podría estar basada en diferentes lenguas, sistemas lingüísticos o vocabularios que sirviera como punto medio entre LO y LM.

Al Ansary (2014: 01) define la interlingua como la creencia de que las estructuras superficiales de las lenguas son muy diferentes entre sí, pero que las estructuras más profundas de las lenguas son comunes a todas o muchas de ellas. Es decir, se asume que es posible convertir textos en LO a representaciones comunes a más de una lengua, también conocidos como universales lingüísticos y de estas representaciones en interlingua, se pueden generar textos en otras lenguas (Abaitua, 2002; Hutchins, 1995: 432).

Según Gavaldà (2000: 02), en los modelos de interlingua se crea primero una interpretación semántica, por ejemplo, «come (niño, pescado)», de una oración en LO, por ejemplo, «el niño come pescado», y luego se crea una versión en LM de dicha interpretación semántica, por ejemplo, «the boy eats fish» en este caso. De esta manera, la traducción se efectuaría en dos pasos: de LO a interlingua (análisis) y de interlingua a LM (generación) (Abaitua, 2002; Hutchins, 1995: 432; Hutchins, 2003: 08). Algunos de los proyectos más importantes en este campo durante esta época fueron ATAMIRI (en Bolivia) y DLT (Distributed Language Translation) de BSO, una compañía de *software*, de Países Bajos y el proyecto Rosetta de Philips en Eindhoven dirigido por Jan Landsbergen (Abaitua, 2002; Hutchins, 1995: 445).

También se retomaron investigaciones que habían comenzado en los años 60 y 70 en el campo de la TA estadística y la TA basada en ejemplos. En 1988, IBM presentó el llamado «Proyecto Candide», un sistema basado en métodos estadísticos, cuyos resultados sorprendieron por su alta calidad y animaron a los investigadores a seguir investigando en este campo. El proyecto se componía de informes en francés y en inglés de los debates parlamentarios canadienses. Este método se basaba en la alineación de oraciones, de grupos de palabras o de palabras individuales en textos paralelos. Tras alinear dichos textos, se calculaba la probabilidad de que ciertas palabras u oraciones del texto en una lengua se correspondieran con otras palabras u

oraciones en el texto alineado en la otra lengua (Berger et al., 1994; Hutchins, 1995: 448; Hutchins, 2001: 23; Ney, 2004: 12; Tertoolen, 2012: 146).

Por otro lado, aunque la TA basada en ejemplos se empezó a introducir en 1984, no fue hasta finales de los años 80 que se empezó a experimentar con ella. La hipótesis en la que se basaba este enfoque es que, en muchas ocasiones, la traducción implica la reutilización de oraciones que se hayan traducido anteriormente de manera similar. De esta manera, el proceso consistía en la selección de oraciones o grupos de palabras equivalentes de un corpus de textos bilingües paralelos, los cuales se han alineado con anterioridad. La principal ventaja que ofrecía este enfoque es que, puesto que las traducciones se han extraído de un corpus de textos previamente traducidos por profesionales, los resultados suelen ser idiomáticos y de calidad (Hutchins, 1995: 448-449; Hutchins, 2001: 23; Hutchins, 2005; Oliver, 2014: 04; Seoane, 2015: 218).

En los años 90, la investigación se empezó a focalizar en el procesamiento del lenguaje natural en el contexto de la inteligencia artificial (IA). No obstante, no se abandonaron los enfoques basados en interlingua (Hutchins, 1995: 449). Además, también surgieron importantes experimentos en sistemas de TA basados en conocimiento o *knowledge-based*, de la mano de la Universidad Carnegie-Mellon en Pittsburg, en la cual se elaboró un prototipo para TA de inglés a japonés y viceversa para traducción de manuales de ordenadores personales (Hutchins, 1995: 447).

En 1992, esta misma Universidad anunció un proyecto colaborativo con la empresa Caterpillar a partir del cual crearían un sistema de TA basado en conocimiento y en lenguaje controlado de gran calidad y a gran escala llamado Catalyst. Este se utilizaría para la TA multilingüe de manuales técnicos de áreas de conocimientos específicas (Hutchins, 1995: 450; Hutchins, 2001: 24; Hutchins, 2005: 03; Hutchins, 2015: 12; Seoane, 2015: 218).

Otros grupos de investigación de la Universidad de Manchester, la Universidad de Bruselas, la Universidad de Grenoble y la Science University de Malasia experimentaron con sistemas de TA basados en dialogo, en que el texto se elabora en un proceso colaborativo entre el humano y la máquina. Además, también creció el interés por la traducción del lenguaje oral, lo que combinaba el reconocimiento de voz y la interpretación (Hutchins, 1995: 451-452).

Al mismo tiempo, el uso de TA fue cada vez mayor en agencias comerciales, gubernamentales y en compañías multinacionales en los años 90 por su necesidad de traducción de grandes volúmenes de documentación técnica. Además, en estos organismos los principales usuarios de estos sistemas no serían necesariamente traductores profesionales. Cada vez se hizo más patente que lo que los traductores profesionales necesitaban eran herramientas que facilitaran

su tarea, por lo que se comenzaron a desarrollar herramientas con acceso a diccionarios, bases de datos terminológicas, glosarios, procesamiento de texto multilingüe, etc. (Hutchins, 1995: 453).

Así, a finales de los años 80 y principios de los 90 reaparece un producto dedicado a facilitar la tarea del traductor profesional que había surgido ya en los años 60, las estaciones de trabajo del traductor o *translator workstations*, como por ejemplo Translation Manager, Trados, Star Transit o EuroLang Optimizer, entre otras. (Hutchins, 1995: 453; Hutchins, 2003: 14; Vargas y Ramírez-Polo, 2011).

Desde el mundo académico, uno de los momentos claves en el desarrollo de las estaciones de trabajo del traductor fue la publicación del artículo de Martin Kay titulado «The Proper Place of Men and Machines in Language Translation», publicado por primera vez en 1980 y, más tarde, en 1997, en la revista «Machine Translation». En dicho artículo, por un lado, se realizaba una crítica a la TA y por otro, se proponían nuevas formas de trabajar para los traductores profesionales (Alonso y Nunes-Vieira, 2017: 345; Vargas y Ramírez-Polo, 2011).

Además, Kay hizo algunas de sus principales aportaciones como, por ejemplo, la propuesta de que los ordenadores podrían asumir de forma gradual algunas de las tareas más repetitivas del proceso de traducción (Vargas y Ramírez-Polo, 2011). Esta aportación de Kay (1997: 13) dice lo siguiente:

I want to advocate an incremental approach to the problem of how machines should be used in language translation. [...] I want to advocate a view of the problem in which machines are gradually, almost imperceptibly, allowed to take over certain functions in the overall translation process. First they will take over functions not essentially related to translation. Then, little by little, they will approach translation itself. The keynote will be modesty. At each stage, we will do only what we know we can do reliably.

Kay (1997) llama *Translator's Amanuensis* al instrumento que debía asistir al traductor en su tarea profesional, es decir, la estación de trabajo o *workstation* del traductor, y este debía estar compuesto por diversas herramientas (Somers, 2003: 13; Vargas y Ramírez-Polo, 2011). Por un lado, debería tener un editor o procesador de texto que mostraría el texto original (TO) en la parte superior de la ventana y el texto meta (TM) en la parte inferior. También debería contar con complementos o extensiones (*add-ons*), como, por ejemplo, diccionarios, bases de datos o un sistema de marcado que resaltara problemas y dificultades en la traducción para que el

traductor pudiera afrontarlas más adelante. Por último, integraría TA con la que el traductor pudiera interactuar (Kay, 1997; Vargas y Ramírez-Polo, 2011).

Hutchins (2003: 14) afirma que muchas de las estaciones de trabajo del traductor empezaron a ofrecer sistemas integrados de traducción totalmente automática como Systran, Logos y Transcend en los años 90. De manera paralela, desde la aparición de la localización aproximadamente en los años 90, la TA se empieza a utilizar también para localizar diversos productos multimedia, en especial *software* y la documentación que lo acompaña, así como la caja, los archivos de ayuda y demás materiales. El uso de la TA en localización estuvo propiciado por la necesidad de que el producto se comercializara en diversos *locales* a la vez, para lo que se requería un gran número de traducciones a diversas lenguas en un espacio muy reducido de tiempo (Hutchins, 2015: 12).

Desde este momento hasta el presente, el mundo de la investigación y el mundo empresarial en este ámbito se han concentrado en desarrollar herramientas de traducción asistida por ordenador o TAO de manera que los traductores puedan desempeñar su profesión de forma más ágil y sencilla, sin que las traducciones pierdan su calidad. A continuación, se observará la evolución de la TA desde los 2000 hasta hoy en día.

#### 2.3.5.5. *Del 2000 hasta la actualidad*

Como se ha mencionado en la sección anterior, los desarrollos llevados a cabo en las estaciones de trabajo del traductor se produjeron por tres factores principales, según Vargas y Ramírez-Polo, (2011). Por un lado, la investigación en TA evidenció la necesidad de herramientas TAO para ayudar a los traductores. Por otro lado, la generalización del uso de los ordenadores y de internet supuso el incremento de la información y la aceleración de los procesos de trabajo, lo que, a su vez, hizo patente la necesidad de automatizar el proceso de traducción. Por último, los avances en el *hardware* informático permitieron la creación de ordenadores muy potentes en tamaños más pequeños (Vargas y Ramírez-Polo, 2011).

Los años 90 y los 2000 estuvieron marcados por la aparición de Internet y la *World Wide Web*, lo que permitió el acceso a gran cantidad de información. En este momento, solo algunas personas podían ser creadoras de contenido y el papel de los usuarios era simplemente el de consumidores de información (Vargas y Ramírez-Polo, 2011). Asimismo, la TA estadística pasó a situarse en el centro de las investigaciones en aquel entonces por 3 razones principales según Hutchins (2005: 04):

- The main reasons are: i) the availability of large monolingual and bilingual corpora,
- ii) the open-source availability of software for performing basic SMT processes



(alignment, filtering, reordering), such as Moses, GIZA, etc.; iii) the availability of widely accepted metrics for evaluating systems (BLEU, and successors).

Es decir, que, gracias a Internet, el acceso a corpus y a información en general se democratizó y se hizo mucho más accesible. Además, aparecieron nuevas aplicaciones de *software* que incorporaban TA estadística y se permitió el acceso a algunas de las principales métricas de evaluación de la calidad de la TA.

Como indica Hutchins (2005: 05), pese a que se produjo un giro hacia enfoques basados en la TA estadística, los investigadores no abandonaron del todo la TA basada en reglas, es más, muchos de ellos adoptaron a partir de este momento enfoques híbridos que combinaban ambos enfoques. Además, el uso de la TA, acompañada de procesos de pre-edición (lenguaje controlado o control de la terminología) y posedición, se generalizó todavía más y se comenzó a utilizar en otros ámbitos de aplicación como las patentes, los subtítulos y la localización, entre otros.

A lo largo de los años 2000, el uso de la TA también cambió tanto por parte de los traductores profesionales como por el público general. Por un lado, los traductores profesionales empezaron a utilizar mecanismos de TA y memorias de traducción (MT) para realizar primeras versiones o borradores de sus traducciones. Por otro lado, los usuarios ya pudieron empezar a acceder a mecanismos de TA en línea gracias a proveedores como Google, con Google Translate, y ya no necesitaban instalar ningún tipo de *software* adicional en sus ordenadores personales (Hutchins, 2005: 05).

En la década de los 90, Kay (1997) sugirió algunas de las herramientas que debía tener un *Translator's Amanuensis* y, en los años 2000, autores como Sommers (2003: 14) comenzaron a defender también la inclusión de procesadores de texto en las estaciones de trabajo de los traductores. Estos procesadores debían llevar incorporadas herramientas como recuento de palabras, correctores ortográficos, diccionarios de sinónimos (tesauros), diccionarios en línea, glosarios, enciclopedias, MT y recursos basados en corpus, como la alineación de fragmentos y la búsqueda de concordancias. Además, poco a poco se empezaron a integrar también otras aplicaciones, como la mensajería instantánea y las *newsletters* (Sommers, 2003: 14; Vargas y Ramírez-Polo, 2011).

Uno de los grandes cambios fue sin duda el paso de la Web 1.0, en que los usuarios eran meros consumidores de contenido, a la Web 2.0, en la cual se promueve la interacción entre usuarios, la redacción de contenido, la interoperabilidad y el diseño centrado en el usuario (Vargas y Ramírez-Polo, 2011). Esto contribuyó a que la información en Internet creciera de forma

exponencial, lo que evidenció la necesidad de que se realizaran las traducciones a gran velocidad, para que estuvieran disponibles en todo el mundo al mismo tiempo, lo que, a su vez, propició el uso de sistemas de TA para la traducción de dichos contenidos.

Así, en 2005, apareció el primer sistema de TA estadística comercial, llamado «Language Weaver». En 2006, este mismo sistema pasó a soportar 27 pares de lenguas. Además, permitía mediante una aplicación incorporada en el sistema, que las empresas que lo adquirían personalizaran el *software* según la terminología utilizada en su campo de trabajo y según su estilo corporativo (Yunker, 2006: 13).

En vista a este nuevo paradigma donde la traducción asistida por ordenador estaba convirtiéndose más bien en TA asistida por traductores humanos profesionales, empezaron a desarrollarse nuevos perfiles dentro de la profesión del traductor, como por ejemplo el predictor y el poseedor, los consultores lingüísticos o los expertos en control de calidad, entre otros, algo que evidenció la necesidad de cambiar la forma en que la traducción se enseñaba en las universidades (Vargas y Ramírez-Polo, 2011).

Como se ha indicado anteriormente, la TA estadística fue el tipo de TA más utilizado y estudiado hasta el momento, pero en la última década, han surgido nuevas técnicas para seguir desarrollando y mejorando los sistemas de TA neuronal, debido a la baja calidad que tenían hasta el momento los textos traducidos mediante TA estadística (Wu et al. 2016). Así, la TA neuronal ha mostrado en estos últimos años un notable progreso con sistemas de vanguardia capaces de producir traducciones de gran calidad, al menos a nivel oracional (Britz et al., 2017: 01; Libovický, Brovelli y Cartoni, 2018: 01; Sánchez-Gijón, Moorkens y Way, 2019: 01; Wu et al., 2016).

Fundamentalmente, los sistemas neuronales se basan en un modelo codificador-decodificador o *encoder-decoder*, en que el modelo codifica la oración en LO y luego la decodifica en LM mediante los principios del aprendizaje automático (*machine learning*) y redes neuronales, recurrentes o convolucionales, que emulan el comportamiento del cerebro mediante complejos algoritmos, para aprender a traducir de manera automática (Bahdanau, Cho y Bengio, 2014; Jean, Cho, Memisevic y Bengio, 2014). Estas redes neuronales se entrenan de manera conjunta mediante corpus bilingües alineados (Castilho et al., 2017: 110; Jean, Cho, Memisevic y Bengio, 2014; Luong, Pham y Manning, 2015). Actualmente, este es el paradigma por excelencia dentro de los estudios en TA y la mayor parte de traductores automáticos de internet, incluidos el Traductor de Google o DeepL, entre otros, utilizan este tipo de sistemas.

Uno de los mayores hitos en la TA y que supuso un gran cambio en la calidad de los textos traducidos mediante TA, fue el lanzamiento en 2016 de la TA neuronal del Traductor de Google. En la última década, el Traductor de Google ha pasado de traducir solo unos pocos idiomas a traducir a alrededor de 133 lenguas, con volúmenes que llegan a más de 140 mil millones de palabras al día (Caswell, 2022; Turovsky, 2016). Además, el sistema produce traducciones con un 60 % menos de errores de traducción en pares de lenguas populares, como, por ejemplo, de inglés a español (Schuster, Johnson y Thorat, 2016; Siddhant et al., 2019: 01; Wu et al., 2016: 20).

Uno de los métodos más novedosos utilizados por la TA neuronal de Google posibilitaba la llamada *zero-shot translation*, un tipo de traducción entre pares de lenguas que nunca se le habían enseñado explícitamente al sistema de TA durante el entrenamiento (Aharoni, Johnson y Firat, 2018: 01; Johnson, 2017; Schuster, Johnson y Thorat, 2016). Por ejemplo, se puede haber entrenado al sistema para traducir de inglés a coreano y viceversa, pero, si se necesitara traducir de coreano a japonés (par de lenguas en el cual no se ha entrenado al sistema), el sistema aplicaría los conocimientos adquiridos en su entrenamiento de inglés a coreano para producir la traducción entre coreano y japonés. En los últimos años, han aparecido otros sistemas de TA neuronal como DeepL, Reverso o Microsoft Translator, entre otros, que también cuentan con mucho potencial.

La TA neuronal se ha introducido como un enfoque con gran capacidad para prevenir y solucionar muchas de las carencias que habían tenido hasta el momento los sistemas tradicionales de TA (Wu et al., 2016: 01). No obstante, hoy en día, la TA todavía no es perfecta y aún presenta algunos problemas, como, por ejemplo, que entrenar estos sistemas requiere de una gran inversión de tiempo y dinero (Britz et al., 2017: 01; Wu et al., 2016: 01). Es más, cuando el sistema se enfrenta a palabras que no aparecen frecuentemente en una lengua o que no se le habían mostrado durante el entrenamiento, la TA neuronal aún produce textos de relativa baja calidad, lo que dificulta su uso para pares de lenguas en que al menos una de las lenguas sea, por ejemplo, muy rica morfológicamente (Ataman, Gangi y Federico, 2018: 31).

Además, sigue habiendo muchas voces críticas sobre la TA neuronal como, por ejemplo, do Carmo (2018: 323), quien afirma que la TA no es un proceso de traducción completo puesto que supone un trabajo adicional por parte del traductor, la posesición. Por su parte, Sánchez-Gijón, Moorkens y Way (2018: 331), opinan que la TA sigue siendo impopular para los traductores profesionales porque su calidad es todavía baja. A esto, Sánchez-Gijón (2016) añade que la TA se ha utilizado en los últimos tiempos para rebajar los costes de traducción en las empresas, lo

que ha perjudicado a los traductores profesionales, puesto que muchos de los textos se traducen mediante TA e incluso se han dejado sin revisar.

Para finalizar el apartado, se deben mencionar también los avances en los campos del reconocimiento de voz, el procesamiento del lenguaje natural y la síntesis de voz, las cuales, junto a la TA representaron, según Gavalda (2000: 01), «las cuatro áreas de mayor impulso en el conjunto de las tecnologías de la lengua y la lingüística computacional», puesto que nos encontramos en la era de la información y las tecnologías de la lengua adquieren una gran relevancia, tanto en el ámbito comercial y profesional como en el investigador y académico.

Además, también han cobrado especial relevancia las disciplinas como la localización, que, aunque ya existía desde hacía décadas, a causa de la creciente internacionalización de los productos y de la globalización de las empresas, se ha convertido en parte esencial del proceso de comercialización de un producto multimedia. No solo se traducen sitios web, sino *software*, videojuegos y, con la aparición de los *smartphones* o teléfonos inteligentes, aplicaciones o *apps*. Se ha profundizado mucho más sobre este tema en el apartado dedicado a la localización en esta tesis.

A modo de conclusión, se podría decir que la TA cuenta con un largo recorrido. Muchas veces se piensa en ella como un sistema muy avanzado y novedoso que tiene la capacidad de traducir los textos casi de manera perfecta e instantánea y que sustituirá a los traductores en un futuro próximo. No obstante, si se analiza su historia desde el principio se observa, por un lado, la gran diversidad de opiniones provenientes de diferentes campos, como son la informática, la lingüística computacional o la traducción. Por otra parte, se percibe la gran complejidad de dichos sistemas, los cuales se han desarrollado a un paso mucho más lento del que muchos vaticinaban. Algo que parece claro, es que la TA, incluso la TA neuronal, no es perfecta todavía, ya que los textos producidos mediante estos sistemas deben ser revisados por traductores humanos profesionales capaces de llevar a cabo un proceso de posesición adecuado que asegure que se mantiene la calidad de las traducciones y que se siguen los principios de la localización.

### 2.3.6. Reflexiones finales sobre TA

Aunque a veces se pueda pensar que la TA es algo novedoso, lo cierto es que esta ya lleva décadas estudiándose. En un principio, los investigadores en esta materia siguieron de cerca los principios de la criptografía y afirmaron de manera optimista que esta sería capaz de reemplazar a los traductores humanos. Esta gran motivación en el estudio de la TA siguió hasta el informe ALPAC, el cual frena en los años 60 el estudio de sistemas de TA, vista la poca probabilidad de

que los sistemas de la época realizaran traducciones completamente automáticas de gran calidad. A partir de este momento, se empieza a progresar en el desarrollo de las herramientas TAO, que asistirían a los traductores profesionales en su profesión y los ayudarían a ser más eficientes, aunque esto no significó abandono total de los estudios en TA.

A lo largo del estudio de la literatura en TA, se ha contemplado el desarrollo de los diferentes modelos existentes. Desde los sistemas basados en reglas, hasta la traducción automática neuronal, la TA se ha ido mejorando y, hoy en día, ocupa una posición relevante dentro del mundo del aprendizaje automático o *machine learning*, más en concreto dentro del ámbito del procesamiento del lenguaje natural o *natural language processing* (NLP). De esta manera, se han identificado algunos de los futuros desafíos de la TAN. En el apartado del funcionamiento de la TAN, se ha observado que los sistemas neuronales traducen a nivel de oración, por eso, uno de los avances que se estudian en la actualidad es la introducción de mejoras para que los sistemas neuronales puedan traducir a nivel de párrafos o incluso textos enteros de manera coherente.

Aunque la TA se perciba (a veces correctamente) como una manera de recortar costes, los traductores no deberían dejarse intimidar (al menos de momento) por aquellos titulares que expresan en mayúsculas y viva voz que se van a quedar sin trabajo. La necesidad de traducciones de calidad sigue existiendo y, además, se han creado nuevos puestos de trabajo para los lingüistas, como la posesición y la localización de webs, *software* o aplicaciones, entre otros productos multimedia, las cuales deberían introducirse en los planes formativos de todos los grados en traducción. Algo que parece estar claro es que la TA está aquí para quedarse, por lo que, aunque los traductores no necesiten conocer todos los aspectos técnicos de los sistemas de TA, es importante tener al menos una visión general de los fundamentos de estos modelos.

### 3. Análisis y resultados

El siguiente apartado presenta la metodología empleada en el análisis llevado a cabo en la presente tesis. Además, se adjuntan algunos extractos de los corpus analizados o de las listas de frecuencias empleadas en los análisis, entre otros materiales de estudio.

Esta sección también presenta los análisis realizados, así como los principales resultados obtenidos de estudiar los corpus de reseñas de hoteles y restaurantes.

#### 3.1. Metodología

Para el análisis de las problemáticas de localización que presentan las reseñas de hoteles y de restaurantes traducidas mediante TA, se han recopilado dos corpus bilingües, uno de hoteles y otro de restaurantes de comida típica valenciana por su alto contenido en vocabulario y aspectos regionales. Los corpus consisten en un total de 500 reseñas escritas originalmente en inglés (250 de restaurantes y 250 de hoteles), extraídas de la plataforma de reseñas Tripadvisor.com que, posteriormente, se han procesado con el motor de TA de la plataforma, el Traductor Google, para conseguir sus traducciones automáticas al español. Posteriormente, se ha analizado el producto de la TA resultante siguiendo las pautas habituales de posesión. Gracias a esto ha sido posible identificar los principales errores que afectan a nivel lingüístico y cultural a la calidad de las reseñas traducidas con TA directamente en la plataforma.

Tanto el análisis como el corpus se han dividido en reseñas de hoteles y en reseñas de restaurantes. Esto se ha hecho así porque cada una de estas reseñas tiene características distintas como, por ejemplo, su longitud, la terminología (con mayor o menor carga cultural) o su estructura, entre otras. Estas características determinarán cómo los motores de TA procesan las reseñas y, por lo tanto, la calidad de los resultados de TAN de los contenidos.

En primer lugar, se ha llevado a cabo un análisis cuantitativo y cualitativo de los dos corpus bilingües de reseñas traducidas al español con TA, es decir, se ha examinado el número de errores de localización en cada uno de los corpus y en qué medida estos afectan a la calidad de las traducciones. En segundo lugar, y como parte del análisis cualitativo, se han estudiado los problemas lingüísticos y culturales de localización, así como aquellas causas frecuentes que afectan al nivel de calidad de las reseñas traducidas con TA, que, en este caso, se trata de errores surgidos por problemas de explicación o duplicidad, problemas por la calidad del motor de TA y problemas que se han producido por un error en el original. Además, se ha realizado una comparativa entre los resultados de los corpus bilingües de reseñas de restaurantes y de hoteles.

En tercer lugar, y para comprender hasta qué punto difieren las reseñas traducidas al español con TAN de las reseñas redactadas por hablantes nativos de español en este idioma, se ha llevado a cabo un análisis de frecuencias para así determinar si un mismo concepto aparece el mismo número de veces en las reseñas escritas originalmente en español que en las reseñas traducidas con TA al español. Para ello, se han recopilado dos corpus monolingües, con un total de 500 reseñas de restaurantes y de hoteles escritas originalmente en español por hablantes nativos de esta lengua, extraídas también de la plataforma Tripadvisor.com. Estos corpus se han procesado con el *software* para análisis de corpus WordSmith y este ha extraído las listas de frecuencias correspondientes. Por otra parte, también se han procesado con WordSmith las reseñas traducidas al español con TA. De esta manera, ha sido posible realizar una comparativa sobre la naturalidad y la autenticidad de las reseñas de los corpus monolingües y las de los corpus bilingües.

Por último, también se ha llevado a cabo un análisis sobre cómo la longitud de las reseñas puede afectar la calidad de los resultados de TA. Para ello, primero se han fijado los parámetros de calidad utilizados en este análisis, establecidos mediante una clasificación propia basada en el modelo DQF de TAUS (Van der Meer et al., 2017), pero en que se ha puesto especial énfasis en las convenciones propias del *locale*, ya que uno de los objetivos era comprobar si la TAN cumple con los principios de la localización. A partir de esta clasificación, se han comprobado las longitudes medias de las reseñas de hoteles y de restaurantes y se ha realizado un análisis para establecer en qué medida la longitud de las reseñas influye en los resultados de TA.

Los resultados de los diferentes análisis reflejan la situación de la TA en las reseñas publicadas por los usuarios en plataformas de Internet como TripAdvisor.com. Así, se confirma la necesidad de implementar una estrategia de posesición correcta en que se localicen los contenidos y se adapten a los *locales* de llegada. Para este análisis se han examinado dos tipos de reseñas, las de hoteles y las de restaurantes. Los errores y problemáticas se han clasificado siguiendo varios criterios, mediante los cuales se han obtenido resultados significativos sobre los problemas que dificultan la localización de ciertas reseñas.

En los apartados siguientes se han incluido fragmentos correspondientes a los corpus utilizados en los diferentes análisis llevados a cabo en la presente tesis. Además, también se han añadido partes de las listas de frecuencias extraídas con WordSmith y de las tablas utilizadas para el análisis de la longitud de las reseñas.

### 3.1.1. Corpus de reseñas bilingüe (Access)

A continuación, se muestran algunos ejemplos de las entradas de la base de datos de Access con que se ha trabajado para analizar el corpus. En concreto, se ha incluido una entrada correspondiente al corpus bilingüe de reseñas de hoteles y otra correspondiente al corpus bilingüe de reseñas de restaurantes. Las entradas se encuentran en la Ilustración 16, 17, 18 y 19, a continuación:

The screenshot shows a web application window titled 'Hoteles'. It displays a bilingual hotel review entry for item number 249. The interface is divided into several sections:

- Observations:** 'no vale el dinero (not worth the money)' in English and 'Visité aquí / hicieron que la habitación fuera muy estrecha para decir lo menos (to say the least)' in Spanish.
- Title:** 'Not worth the money' in English and 'No vale el dinero' in Spanish.
- Review Text:** A detailed English review on the left and its Spanish translation on the right.
- Filters:** A list of checkboxes on the right side, including 'revisado carmen', 'cultural', 'lingüístico', 'error en original', and 'problema calidad TA'. The 'lingüístico' filter is checked.
- Quality:** A dropdown menu labeled 'CALIDAD TA' set to 'Aceptable'.
- Comments:** A text area labeled 'comentarios para paper' with the text 'Tú y usted'.

Ilustración 16. Ejemplo de entrada de reseña de hotel.

The screenshot shows a web application window titled 'Hoteles'. It displays a bilingual hotel review entry for item number 4. The interface is divided into several sections:

- Observations:** 'Un doble (una doble) / cama masiva /Maritin Serreria / Tenemos un desayuno / Unas tapas encantadoras / el costo general de la comida' in English and 'Nos alojamos por 3 noches / obtener un poco de sol /articulos de aseo encantadores /De nuevo tan fácil.' in Spanish.
- Title:** 'Perfect' in English and 'Perfecta(femenino) //Perfecto(masculino)' in Spanish.
- Review Text:** A detailed English review on the left and its Spanish translation on the right.
- Filters:** A list of checkboxes on the right side, including 'revisado carmen', 'cultural', 'lingüístico', 'error en original', 'problema calidad TA', and 'problema explicaciones/duplicidade'. The 'revisado carmen' and 'lingüístico' filters are checked.
- Quality:** A dropdown menu labeled 'CALIDAD TA' set to 'Inaceptable'.
- Comments:** A text area labeled 'comentarios para paper' with the text 'masculinos y femeninos de los títulos / ella (nos) preguntó (omisión pronombres)'.

Ilustración 17. Ejemplo de entrada de reseña de hotel.



restaurantes		<input type="button" value="←"/> <input type="button" value="→"/>		<input checked="" type="checkbox"/> revisado carmen
núm item	251	más observaciones	la mejor comida que hemos tenido en Valencia / teníamos dos vegetarianos con nosotros	<input type="checkbox"/> cultural
obsevaciones	Stop by for the best meal / Pasa por la mejor comida	TA título	DEBES VERLO	<input type="checkbox"/> lingüístico
paper 2		TA a ES	Este restaurante fue absolutamente increíble. El propietario, Faud, estaba muy bien informado y era muy servicial. Recomendamos las tapas que deberíamos probar y es, con mucho, la mejor comida que hemos tenido en Valencia. Teníamos dos vegetarianos con nosotros y fue extremadamente servicial. Este lugar es de visita obligada. Pasa por la mejor comida	<input type="checkbox"/> error en original
Título	A MUST SEE			<input checked="" type="checkbox"/> problema calidad TA
Reseña	This restaurant was absolutely amazing. The owner, Faud, was extremely knowledgeable and very accommodating. We recommended the tapas we should try and it's by far the best meal we've had in Valencia. We had two vegetarians with us and he was extremely accommodating. This place is a must see!! Stop by for the best meal			<input type="checkbox"/> problema explicaciones/duplicidade
				CALIDAD TA <input type="text" value="Aceptable"/>
				comentarios para paper <input type="text" value="C: servicial (accommodating)"/>

Ilustración 18. Ejemplo de entrada de reseña de restaurante.

restaurantes		<input type="button" value="←"/> <input type="button" value="→"/>		<input checked="" type="checkbox"/> revisado carmen
núm item	242	más observaciones	servicial y acogedor	<input type="checkbox"/> cultural
obsevaciones	a la guindilla	TA título	Algo especial	<input type="checkbox"/> lingüístico
paper 2		TA a ES	Nos obsequiaron con unas tapas exquisitas. Gambas al ajillo y a la guindilla perfectamente elaboradas y cocidas, calamares a la plancha, anchoas marinadas, croquetas mixtas. El anfitrión fue muy servicial y acogedor. Qué suerte estar a la vuelta de la esquina, así que volveremos más tarde en nuestro viaje a Valencia.	<input type="checkbox"/> error en original
Título	Something special			<input checked="" type="checkbox"/> problema calidad TA
Reseña	We were treated to some exquisite tapas. Perfectly prepared and cooked garlic and chilli prawns, grilled squid, marinated anchovies, mixed croquettes. The host was very helpful and welcoming. So lucky to be staying just around the corner, so will be returning later in our trip to Valencia.			<input type="checkbox"/> problema explicaciones/duplicidade
				CALIDAD TA <input type="text" value="Aceptable"/>
				comentarios para paper <input type="text"/>

Ilustración 19. Ejemplo de entrada de reseña de restaurante.

Como se puede observar, los campos que se incluyen en el corpus son el número que ocupa la reseña dentro del corpus, algunos campos para incluir observaciones y comentarios, las reseñas y sus títulos en inglés escritas en paralelo junto a sus traducciones automáticas, así como casillas para marcar los tipos de problemas encontrados en cada reseña y un desplegable para indicar el nivel de calidad de estas, entre otros

### 3.1.2. Corpus monolingüe (Excel)

A continuación, se ha incluido un extracto del corpus monolingüe de reseñas escritas originalmente en español de España.

En dicho corpus se han incluido tanto el título como el cuerpo de la reseña, además del nombre de usuario de los autores de las reseñas ya que este puede aportar información importante de la nacionalidad o procedencia de los autores. Además de eso, se ha incluido la calificación general que han dado dichos usuarios a los hoteles o restaurantes valorados, ya que esto también permite detectar patrones de valoración entre los usuarios de los diferentes *locales* para determinar si estos califican los productos o servicios de forma diferente. Por este mismo motivo, también se ha incluido una columna de valoraciones de servicios concretos.

Tabla 7. Corpus monolingüe (Excel).

<b>Título</b>	<b>Reseña</b>	<b>Autor</b>	<b>Calificación General(1-5)</b>	<b>Valoraciones concretos</b>	<b>Servicios</b>
<b>Buen hotel en el centro de Valencia</b>	Hotel situado en el centro de Valencia. Tiene parking con acceso directo al hotel. Pero lo mejor es la atención de todos los empleados. Solo buscan la satisfacción total del cliente. Recomiendo este hotel para todo tipo de personas.	E6484HKjosep	5	Sí	
<b>Céntrico</b>	Hotel en pleno centro de Valencia. Hay bastantes restaurantes en las misma calle y en las de alrededor. Tienes muchos monumentos a los que acudir a pie. Hay una estación cerca, por lo que	Ainho	5	No	

	<p>moveirse por el resto de la comunidad es bastante práctico. 7,9/10</p>			
<b>Buena experiencia.</b>	<p>Hotel céntrico en Valencia. Con todas las medidas de seguridad por la COVID 19. Habitación amplia con baño con todas las comodidades esperadas. Zona tranquila con un acceso a todos los sitios importantes de manera peatonal. Aparcamiento bajo el hotel aunque es privado.</p>	Jose B	5	Sí
<b>Excelente</b>	<p>Hotel inmejorable, con una situación excelente, cerca de bus, metro y estación de tren. El personal muy amable y dispuesto para ayudar. El desayuno buffet rico y variado y con todas las medidas de seguridad.Lo recomendamos totalmente.</p>	Maku	5	-
<b>Un hotel con encanto</b>	<p>Buena ubicación, justo en el centro mismo, con todas las comodidades y medidas de seguridad por el Covid-19. Personal muy atento y muy amable, desde las camareras de piso, el personal de recepción y camareros de salón. La habitación amplia, cómoda y muy limpia... Un hotel a repetir sin duda</p>	laury b	5	-

<b>MUY RECOMENDABLE</b>	Como todos los Vincci, recomendable. Lugar acogedor, servicio de desayuno perfecto, atención exquisito, parking y en el mismo centro de Valencia. la coordinación parking-recepción a veces flojea pero con un poco de paciencia se soluciona.	170antoniomanuelh	4	Sí
<b>Muy bien</b>	El hotel nos ha gustado mucho. En mi opinión el plus que ofrece es el parking, pues es un parking público pero con plazas reservadas para el hotel. Encontrar un hotel céntrico en Valencia y con parking es complicado. Sólo quiero indicar que como hay que entrar al parking por una calle peatonal es fácil pasarselo. Lo mejor Es poner en el navegador calle Pascual y Genis, y desde esa calle se ve ya la entrada al parking lys.	montseo224	5	-

### 3.1.3. Listas de frecuencias de WordSmith

Las siguientes tablas muestran una parte de las listas de frecuencias extraídas con WordSmith durante el análisis llevado a cabo en esta tesis.

Se han extraído las siguientes listas de frecuencias: 1) una lista de frecuencias de las palabras encontradas en el corpus de reseñas de hoteles en el corpus de ES-TA, 2) una lista de frecuencias de las palabras encontradas en el corpus de reseñas de hoteles en el corpus de ES Original, 3) una lista de frecuencias de las palabras encontradas en las reseñas de restaurantes en el corpus de ES-TA y 4) una lista de frecuencias de las palabras encontradas en el corpus de reseñas de restaurantes en el corpus de ES Original.

*Tabla 8. Lista de frecuencias de palabras encontradas en las reseñas de hoteles (Corpus ES-TA).*

<b>Word</b>	<b>Freq.</b>	<b>%</b>	<b>Texts</b>	<b>%2</b>	<b>Dispersion</b>
<b>HOTEL</b>	423	1,57	1	100,00	0,93
<b>HABITACIÓN</b>	199	0,74	1	100,00	0,90
<b>PERSONAL</b>	191	0,71	1	100,00	0,91
<b>DESAYUNO</b>	169	0,63	1	100,00	0,83
<b>EXCELENTE</b>	123	0,46	1	100,00	0,83
<b>HABITACIONES</b>	114	0,42	1	100,00	0,91
<b>CIUDAD</b>	103	0,38	1	100,00	0,86
<b>AMABLE</b>	98	0,36	1	100,00	0,85
<b>UBICACIÓN</b>	98	0,36	1	100,00	0,78
<b>VALENCIA</b>	93	0,34	1	100,00	0,86
<b>PLAYA</b>	86	0,32	1	100,00	0,67
<b>PISCINA</b>	84	0,31	1	100,00	0,56
<b>SERVICIO</b>	82	0,30	1	100,00	0,83
<b>RESTAURANTE</b>	76	0,28	1	100,00	0,77
<b>SERVICIAL</b>	65	0,24	1	100,00	0,80
<b>CERCA</b>	64	0,24	1	100,00	0,77
<b>AGRADABLE</b>	63	0,23	1	100,00	0,90
<b>SOLO</b>	61	0,23	1	100,00	0,90
<b>BUENA</b>	60	0,22	1	100,00	0,81
<b>GRAN</b>	60	0,22	1	100,00	0,86
<b>COMIDA</b>	59	0,22	1	100,00	0,87
<b>CENTRO</b>	54	0,20	1	100,00	0,78
<b>POCO</b>	54	0,20	1	100,00	0,85
<b>NOCHE</b>	52	0,19	1	100,00	0,87
<b>BUEN</b>	51	0,19	1	100,00	0,78
<b>DÍA</b>	51	0,19	1	100,00	0,90
<b>MINUTOS</b>	50	0,19	1	100,00	0,78
<b>BAR</b>	49	0,18	1	100,00	0,86

<b>BUENO</b>	48	0,18	1	100,00	0,84
<b>RECEPCIÓN</b>	48	0,18	1	100,00	0,84
<b>CAMA</b>	47	0,17	1	100,00	0,70
<b>RESTAURANTES</b>	47	0,17	1	100,00	0,80
<b>ESTRELLAS</b>	43	0,16	1	100,00	0,81
<b>REALMENTE</b>	41	0,15	1	100,00	0,83
<b>NOCHES</b>	39	0,14	1	100,00	0,83
<b>LIMPIO</b>	38	0,14	1	100,00	0,84
<b>BAÑO</b>	33	0,12	1	100,00	0,83
<b>CÓMODA</b>	33	0,12	1	100,00	0,72
<b>DEFINITIVAMENTE</b>	32	0,12	1	100,00	0,87
<b>MAR</b>	32	0,12	1	100,00	0,59
<b>PRECIO</b>	32	0,12	1	100,00	0,72
<b>DÍAS</b>	31	0,11	1	100,00	0,88
<b>ESTANCIA</b>	31	0,11	1	100,00	0,80
<b>PROBLEMA</b>	31	0,11	1	100,00	0,86
<b>TENÍAMOS</b>	31	0,11	1	100,00	0,87
<b>ÁREA</b>	30	0,11	1	100,00	0,66

Tabla 9. Lista de frecuencias de palabras encontradas en las reseñas de hoteles (Corpus ES Original).

<b>Word</b>	<b>Freq.</b>	<b>%</b>	<b>Texts</b>	<b>%2</b>	<b>Dispersion</b>
<b>HOTEL</b>	297	1,72	1	100,00	0,92
<b>HABITACIÓN</b>	134	0,78	1	100,00	0,87
<b>PERSONAL</b>	115	0,67	1	100,00	0,90
<b>DESAYUNO</b>	90	0,52	1	100,00	0,88
<b>VALENCIA</b>	71	0,41	1	100,00	0,80
<b>BUENA</b>	62	0,36	1	100,00	0,77
<b>HABITACIONES</b>	61	0,35	1	100,00	0,83
<b>CAMA</b>	58	0,34	1	100,00	0,73
<b>SERVICIO</b>	48	0,28	1	100,00	0,92
<b>AMABLE</b>	44	0,26	1	100,00	0,85
<b>PRECIO</b>	41	0,24	1	100,00	0,76
<b>UBICACIÓN</b>	41	0,24	1	100,00	0,79
<b>CALIDAD</b>	40	0,23	1	100,00	0,85
<b>CENTRO</b>	40	0,23	1	100,00	0,72
<b>RECEPCIÓN</b>	40	0,23	1	100,00	0,86
<b>BAÑO</b>	38	0,22	1	100,00	0,88
<b>EXCELENTE</b>	38	0,22	1	100,00	0,87
<b>LIMPIEZA</b>	38	0,22	1	100,00	0,89
<b>ATENCIÓN</b>	36	0,21	1	100,00	0,81
<b>BUEN</b>	36	0,21	1	100,00	0,64
<b>DUDA</b>	36	0,21	1	100,00	0,77
<b>ZONA</b>	36	0,21	1	100,00	0,83
<b>LIMPIO</b>	35	0,20	1	100,00	0,71

CERCA	33	0,19	1	100,00	0,74
MEJOR	32	0,19	1	100,00	0,76
MUCHO	32	0,19	1	100,00	0,75
PLAYA	32	0,19	1	100,00	0,72
PISCINA	31	0,18	1	100,00	0,54
PARKING	30	0,17	1	100,00	0,53
INSTALACIONES	29	0,17	1	100,00	0,69
TRATO	29	0,17	1	100,00	0,83
GRANDE	28	0,16	1	100,00	0,71
RESTAURANTE	26	0,15	1	100,00	0,68
ESTANCIA	25	0,15	1	100,00	0,80
ESTRELLAS	25	0,15	1	100,00	0,73
FIN	25	0,15	1	100,00	0,82
NOCHE	25	0,15	1	100,00	0,78
SPA	25	0,15	1	100,00	0,57
BUENO	24	0,14	1	100,00	0,85
CIUDAD	24	0,14	1	100,00	0,72
AGRADABLE	23	0,13	1	100,00	0,76
COVID	23	0,13	1	100,00	0,59
GRAN	23	0,13	1	100,00	0,76
GUSTO	23	0,13	1	100,00	0,80
MAR	23	0,13	1	100,00	0,61
RECOMENDABLE	23	0,13	1	100,00	0,66

Tabla 10. Lista de frecuencias de palabras encontradas en las reseñas de restaurantes (ES-TA).

Word	Freq.	%	Texts	%2	Dispersion
COMIDA	220	1,21	1	100,00	0,90
RESTAURANTE	140	0,77	1	100,00	0,90
VALENCIA	107	0,59	1	100,00	0,92
PROPIETARIO	97	0,53	1	100,00	0,90
EXCELENTE	89	0,49	1	100,00	0,90
SERVICIO	82	0,45	1	100,00	0,93
PLATOS	81	0,44	1	100,00	0,89
TAPAS	80	0,44	1	100,00	0,84
VINO	76	0,42	1	100,00	0,88
GRAN	74	0,41	1	100,00	0,90
PAELLA	73	0,40	1	100,00	0,83
AMABLE	61	0,33	1	100,00	0,86
MEJOR	59	0,32	1	100,00	0,91
PERSONAL	58	0,32	1	100,00	0,79
INCREÍBLE	51	0,28	1	100,00	0,83
BUENA	46	0,25	1	100,00	0,91
TUVIMOS	46	0,25	1	100,00	0,87

CROQUETAS	45	0,25	1	100,00	0,86
ANFITRIÓN	44	0,24	1	100,00	0,88
AGRADABLE	42	0,23	1	100,00	0,85
EXPERIENCIA	41	0,22	1	100,00	0,79
MENÚ	41	0,22	1	100,00	0,90
DELICIOSA	39	0,21	1	100,00	0,75
DUEÑO	39	0,21	1	100,00	0,79
TENÍAMOS	39	0,21	1	100,00	0,67
CALIDAD	37	0,20	1	100,00	0,85
AMBIENTE	35	0,19	1	100,00	0,83
PEDERNIZ	35	0,19	1	100,00	0,81
VINOS	35	0,19	1	100,00	0,84
NOCHE	34	0,19	1	100,00	0,78
MEJORES	33	0,18	1	100,00	0,71
PLATO	33	0,18	1	100,00	0,80
BUEN	31	0,17	1	100,00	0,79
REALMENTE	31	0,17	1	100,00	0,77
CASA	30	0,16	1	100,00	0,86
DÍA	30	0,16	1	100,00	0,88
ARROZ	29	0,16	1	100,00	0,54
POCO	29	0,16	1	100,00	0,73
PROBAR	28	0,15	1	100,00	0,75
RECOMIENDO	28	0,15	1	100,00	0,78
DEFINITIVAMENTE	27	0,15	1	100,00	0,81
RECOMENDACIONES	26	0,14	1	100,00	0,86
BUENO	25	0,14	1	100,00	0,86
DELICIOSO	25	0,14	1	100,00	0,82
MARISCOS	23	0,13	1	100,00	0,79
POSTRE	23	0,13	1	100,00	0,79

Tabla 11. Lista de frecuencias de palabras encontradas en las reseñas de hoteles (Corpus ES Original).

Word	Freq.	%	Texts	%2	Dispersion
ARROZ	159	1,26	1	100,00	0,94
PAELLA	83	0,66	1	100,00	0,92
SERVICIO	78	0,62	1	100,00	0,90
COMIDA	73	0,58	1	100,00	0,88
CALIDAD	67	0,53	1	100,00	0,82
PRECIO	64	0,51	1	100,00	0,88
BUENA	61	0,48	1	100,00	0,84
RESTAURANTE	54	0,43	1	100,00	0,89
CARABINEROS	50	0,40	1	100,00	0,69
TRATO	50	0,40	1	100,00	0,84
BUEN	49	0,39	1	100,00	0,86



<b>ENTRANTES</b>	48	0,38	1	100,00	0,86
<b>MEJOR</b>	45	0,36	1	100,00	0,83
<b>PEDIMOS</b>	45	0,36	1	100,00	0,85
<b>EXCELENTE</b>	42	0,33	1	100,00	0,87
<b>ATENCIÓN</b>	41	0,33	1	100,00	0,82
<b>COMER</b>	41	0,33	1	100,00	0,89
<b>ARROCES</b>	40	0,32	1	100,00	0,75
<b>PALMAR</b>	40	0,32	1	100,00	0,81
<b>SITIO</b>	40	0,32	1	100,00	0,78
<b>RECOMENDABLE</b>	38	0,30	1	100,00	0,81
<b>BUENO</b>	37	0,29	1	100,00	0,83
<b>MELOSO</b>	37	0,29	1	100,00	0,75
<b>ESPECTACULAR</b>	36	0,29	1	100,00	0,88
<b>VOLVEREMOS</b>	36	0,29	1	100,00	0,78
<b>POSTRES</b>	29	0,23	1	100,00	0,81
<b>ZONA</b>	29	0,23	1	100,00	0,88
<b>DUDA</b>	28	0,22	1	100,00	0,74
<b>FUIMOS</b>	28	0,22	1	100,00	0,85
<b>SABOR</b>	28	0,22	1	100,00	0,68
<b>SENYORET</b>	28	0,22	1	100,00	0,83
<b>PEBRE</b>	27	0,21	1	100,00	0,73
<b>PUNTO</b>	27	0,21	1	100,00	0,74
<b>ALL</b>	24	0,19	1	100,00	0,68
<b>LOCAL</b>	24	0,19	1	100,00	0,87
<b>COMIMOS</b>	23	0,18	1	100,00	0,83
<b>MUCHO</b>	23	0,18	1	100,00	0,83
<b>PLATO</b>	23	0,18	1	100,00	0,86
<b>TOMATE</b>	23	0,18	1	100,00	0,85
<b>ACOGEDOR</b>	22	0,17	1	100,00	0,69
<b>AMABLE</b>	22	0,17	1	100,00	0,77
<b>COCINA</b>	21	0,17	1	100,00	0,72
<b>FAMILIA</b>	21	0,17	1	100,00	0,90
<b>VALENCIA</b>	21	0,17	1	100,00	0,84
<b>AGRADABLE</b>	20	0,16	1	100,00	0,87
<b>CLOCHINAS</b>	20	0,16	1	100,00	0,87

### 3.1.4. Tablas para el análisis de la longitud de las reseñas

A continuación (Tabla 12 y 13), se presentan extractos de las tablas utilizadas para el análisis de cómo afecta la longitud de las reseñas a la calidad de los resultados de TA.

Tabla 12. Tabla sobre la longitud de las reseñas de hoteles en todos los corpus.

Original EN	Nº palabras EN	ES-TA	Nº palabras TA	CALIDAD TA	Original ES	Nº Palabras
I booked this supposedly 5* property back i	973	Reservé esta supuesta propiedad de 5 * en	1726	Inaceptable	Siempre que doy esta puntuac	345
Upon arrival at this property we upgraded p	1031	A su llegada a esta propiedad, mejoramos p	968	Inaceptable	Hemos estado francamente a g	665
I was among a team of people from my cor	886	Estuve entre un equipo de personas de mi o	914	Inaceptable	Antes que nada quiero decir qu	312
Working in the travel industry for more tha	700	Trabajando en la industria de viajes durante	787	Inaceptable	Ante todo decirles que la estar	234
We just returned from 6 nights at Las Arena	664	Acabamos de regresar de 6 noches en Las A	778	Inaceptable	El hotel es coqueto y con una d	223
Hi, I have booked 10 nights at las arenas.2 cc	627	Hola, he reservado 10 noches en las arenas.	599	Inaceptable	Elegí este hotel para regalar un	197
The room was nicely decorated and clean a	400	La habitación estaba muy bien decorada y li	463	Aceptable	apareció una cucaracha de gran	182
We stayed in Las Arenas for a couple of nigh	457	Nos alojamos en Las Arenas por un par de n	439	Inaceptable	Llevábamos mucho tiempo cor	187
We stayed in Vincci Lys twice within one w	422	Nos alojamos en Vincci Lys dos veces en un	386	Aceptable	Hotel situado cerca de la ciuda	182
The staff at this hotel were incredibly welc	249	El personal de este hotel fue increíblement	369	Inaceptable	Para ser un GL han habido deta	148
A lovely hotel with (mostly) very good atte	333	Un hotel encantador con (en su mayoría) m	364	Aceptable	Me he alojado 7 veces en este	169
We had a lovely stay at Vincci Lys. We staye	352	Tuvimos una estancia encantadora en Vinc	357	Inaceptable	La atención del personal es cor	174
We just spent 5 nights here with our three l	343	Acabamos de pasar 5 noches aquí con nuest	333	Aceptable	El hotel en sí es una pasada. Tie	173
Fantastic location for this lovely hotel whic	294	Fantástica ubicación para este encantador h	297	Inaceptable	Magnifico hotel . Excelente ub	166
My husband and I enjoyed a three day wee	287	Mi marido y yo disfrutamos de un fin de ser	295	Aceptable	Es la segunda vez que nos hosp	148
The location is convenient to the train stati	292	La ubicación es conveniente para la estació	290	Inaceptable	A/ se trata de un hotel de 5 est	154
Only 2 hotels for a deluxe stay in Valencia:	231	Solo 2 hoteles para una estancia de lujo en	267	Inaceptable	He llegado hoy ,hace un mes e	142
We have just returned from 3 amazing nigh	237	Acabamos de regresar de 3 noches increíb	249	Inaceptable	Hotel espectacular. Merece el (	144

Tabla 13. Tabla sobre la longitud de las reseñas de restaurantes en todos los corpus.

Original EN	Nº Palabras EN	TA a ES	Nº Palabras TA	CALIDAD	Original ES	Nº Palabr
We started this tremendous meal with a ni	339	Comenzamos esta tremenda comida co	364	Aceptable	Hoy he comido en La Cambra con suegros,	192
Close to the Joaquin train station in Valenc	346	Cerca de la estación de tren Joaquín de	363	Aceptable	Creo, que si al arroz pudieran preguntarle	247
El Pederniz is located in a residential area	271	El Pederniz está ubicado en una zona re	281	Aceptable	Estuvimos por casualidad y ha sido una op	182
I've been there twice now and both times	248	He estado allí dos veces y las dos veces	253	Aceptable	Hoy hemos visitado este restaurante del F	178
We stayed for a month in Valencia, eating c	233	Nos quedamos un mes en Valencia, cer	241	Aceptable	El local, sencillo, modesto, algo pasado y a	167
This is a very small typically simple Spanis	227	Este es un restaurante español muy pe	231	Inaceptable	Hicimos una visita en familia un viernes a	152
We didn't know about this place, but were	232	No sabíamos sobre este lugar, pero otro	221	Inaceptable	Restaurante muy recomendable para prok	156
My wife and I were stopping through Vale	205	Mi esposa y yo hicimos una parada en V	202	Aceptable	Leímos lo generosos que eran los arroces y	148
I was taken here with work colleagues by c	199	Me llevó aquí con compañeros de traba	190	Inaceptable	Fuimos de excursión a la albufera y buscál	132
Everything about this place was superb - th	178	Todo acerca de este lugar fue excelente	178	Inaceptable	Vinimos atraídos por el plato Michelin que	128
It is difficult to find a good restaurant in Va	176	Es difícil encontrar un buen restaurante	169	Aceptable	Reserva para 2 el mismo día por la mañana	123
Never have I been to a restaurant and had	177	Nunca había estado en un restaurante y	168	Correcta	Hemos pedido mesa bastante tarde y esta	126
A surprising discovery in a quiet area but w	162	Un descubrimiento sorprendente en ur	166	Aceptable	La paella es el típico plato valenciano... Te	119
Even though we were staying around the c	161	A pesar de que nos estábamos quedand	166	Correcta	Espectacular. Un lugar pequeñito, tranqui	111
Traditional setting, authentic food, genial	155	Ambiente tradicional, comida auténtica	165	Aceptable	Decidimos ir a este local a través de Tripac	108
We have been to El Pederniz twice and pla	162	Hemos estado dos veces en El Pederniz	164	Aceptable	Entre ese exceso de oferta hostelera que	116
Nice rustic restaurant a short walk from me	157	Bonito restaurante rústico a pocos pasc	163	Inaceptable	No es fácil salir a comer, buscar opciones v	110
My friends and I were in Valencia to run th	156	Mis amigos y yo estábamos en Valencia	149	Aceptable	Fui con muchas expectativas, y no me defi	110

### 3.2. Análisis cuantitativo y cualitativo del corpus de reseñas sobre restaurantes traducidas con TA

En primer lugar, se ha llevado a cabo un análisis de los datos generales sobre localización, es decir, se han examinado los errores culturales y los errores lingüísticos que han surgido en el corpus de restaurantes. Los errores culturales hacen referencia a aquellas palabras o conceptos propios de la cultura de cierto *locale*, en este caso, el español y, en concreto, el valenciano, que la TAN ha traducido incorrectamente y que deberían localizarse adecuadamente. Por otra parte, con errores lingüísticos se alude a los errores cometidos por la TAN en cuanto a la lengua en sí, o los que comúnmente se estudian en la fase de posesición, es decir, estructuras gramaticales o cambios de significados, entre otros.

#### 3.2.1. Análisis cuantitativo de aspectos de localización en el corpus de reseñas de restaurantes

Para el presente análisis, se ha calculado el porcentaje que representan estos errores con respecto al número total de reseñas, así como el número de reseñas presentes en el corpus con este tipo de errores. La Tabla 14 y la Ilustración 17, a continuación, muestran los resultados correspondientes al corpus de reseñas de restaurantes:

Tabla 14. Datos sobre número de reseñas sobre restaurantes y los principales tipos de errores de localización.

<b>Total de reseñas</b>	<b>250</b>
<b>Nº reseñas con errores culturales</b>	<b>44</b>
<b>Nº reseñas con errores lingüísticos</b>	<b>108</b>
<b>Total de errores de localización</b>	<b>152</b>

El análisis del corpus de restaurantes, con un total de 250 reseñas, ha dado como resultado un total de 44 errores de localización de tipo cultural y un total de 108 errores de localización de tipo lingüístico. Esto es un total de 152 casos de errores de localización. Esto se observa claramente en la Ilustración 20, a continuación, donde se ofrecen los porcentajes que representan estas reseñas con errores de localización con respecto al total de reseñas del corpus de restaurantes:

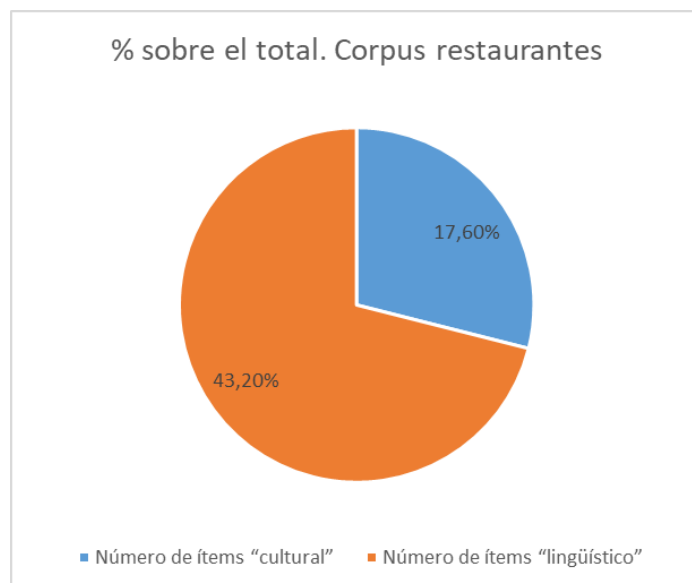


Ilustración 20. Datos sobre porcentaje de reseñas sobre restaurantes y los principales tipos de errores de localización.

La Ilustración 20 ofrece algunos de los principales datos del corpus en referencia a las reseñas de restaurantes. Del total de reseñas de este corpus, 108 contienen errores lingüísticos de localización, es decir, el 43,2 %. En relación con el desempeño de la TA cuando se traducen al español las reseñas de este corpus, se ha observado que esta siempre se equivoca en el uso del verbo «tener» en relación con la comida cuando en inglés se utiliza el verbo *to have*. Se pueden observar ejemplos de este tipo de errores en la Tabla 15 siguiente:

Tabla 15. Errores con la traducción del verbo *to have* por «tener».

Original	TA
<i>We <u>had</u> paella and wine here</i>	Teníamos paella y vino aquí
<i>Meals i've <u>had</u></i>	Comidas que he <u>tenido</u>
<i>I could <u>have</u> three portion as starter</i>	Podría <u>tener</u> tres porciones como entrante principal
<i>We had a leisurely lunch</i>	<u>Tener</u> un almuerzo relajado
<i>I <u>had</u> coffee</i>	<u>Tuve</u> un café
<i>The best food we <u>have had</u> in Spain</i>	La mejor comida que <u>hemos tenido</u> en España
<i>Fresh food you can <u>have</u> in Valencia</i>	Más fresca comida que puedas tener en Valencia
<i>Managed to <u>have</u> 2 excellent dishes</i>	Logramos <u>tener</u> 2 platos excelentes
<i>Best food we <u>have had</u> while in Valencia</i>	La mejor comida que <u>hemos tenido</u> en Valencia
<i>You want to <u>have</u> a very tasteful diner</i>	Quieres <u>tener</u> una cena con mucho gusto

De manera similar, la TA confunde los usos de los verbos «ser» y «estar», sobre todo en relación con la comida, cuando en inglés aparece el verbo *to be* (ver Tabla 16):

Tabla 16. Errores con la traducción del verbo to be por «ser/estar».

Original	TA
<i>Paella <u>was</u> excellent</i>	Paella <u>era</u> excelente
<i>Paella <u>was</u> incredible</i>	La paella <u>fue</u> excelente
<i>All our food <u>was</u> excellent</i>	Toda nuestra comida <u>era</u> excelente
<i>Potatoes <u>were</u> excellent</i>	La papas <u>eran</u> excelentes
<i>Our paella <u>was</u> very tasty</i>	Nuestra paella <u>era</u> muy sabrosa
<i>Food <u>was</u> amazing delicious</i>	La comida <u>era</u> increíblemente deliciosa
<i>The meal <u>was</u> great</i>	La comida <u>era</u> estupenda
<i>Food quality <u>was</u> fantastic</i>	La comida <u>era</u> fantástica
<i>Flan and pineapple carpaccio, which <u>were</u> also delicious</i>	Flan y piña, que también <u>eran</u> deliciosos
<i>The aircon <u>was</u> perfect</i>	El aire acondicionado <u>era</u> perfecto

En los ejemplos de la Tabla 16, se aprecia como la TAN ha utilizado el verbo «ser» cuando, en estos casos, lo correcto hubiera sido utilizar el verbo «estar». En algunos de estos casos incluso se debería haber reformulado la oración, como es el caso de «El aire acondicionado era perfecto», cuya traducción debería haberse reformulado para crear una oración como «El aire acondicionado estaba a una temperatura perfecta».

En cuanto a las expresiones idiomáticas originarias del inglés, la TAN realiza una traducción literal en la mayoría de los casos que da como resultado un texto que no suena nativo para los hispanohablantes (Tabla 17).

Tabla 17. Errores en la traducción de expresiones idiomáticas.

Original	TA
<i>We did our research and found out that</i>	Hicimos nuestra investigación y encontramos que
<i>Was out of this world</i>	Estaba fuera de este mundo
<i>Then the paella - first class</i>	Luego la paella - primera clase
<i>Was off the chain good</i>	Estaba fuera de la cadena bueno
<i>Thus spake a vegetarian</i>	Por lo que habló un vegetariano
<i>A ton of food</i>	Una tonelada de comida
<i>As good as it gets</i>	Tan Bueno como se pone
<i>Nice touch</i>	Buen toque
<i>Small tips are <u>the order of the day</u> in Spain.</i>	Las pequeñas propinas <u>están a la orden del día</u> en España
<i>Didn't even bat an lid</i>	Ni siquiera cerró la tapa

Por otro lado, en el corpus se han identificado 44 reseñas con errores de localización de tipo cultural, es decir, el 17,6 %. Los errores más comunes en este caso han sido el uso de palabras de variantes de español de Latinoamérica, que se han considerado erróneas porque en la presente tesis se toma como referencia el *locale* hablante de español de España. Además, la TA también ha cometido errores en la traducción de los nombres de lugares o denominaciones sobre todo con «l'Albufera» (que ha traducido de manera incorrecta como «la laguna», «Albufera», «L Albufera») o «una ciudad pequeña» hablando del pueblo valenciano de «El Palmar», una pequeña población en las cercanías de la ciudad de Valencia, rodeada por l'Albufera. Algunos de estos errores se pueden apreciar en la Tabla 18, a continuación:

Tabla 18. Errores de localización de tipo cultural.

Original	TA	Comentario
<i>Potatoes</i>	Papas	Aparece en un elevado número de ocasiones. Variante latinoamericana, en España se dice «patatas»
<i>Chilli and garlic prawns</i>	Gambas con ají y al ajillo	Variante latinoamericana, en ES «gambas al ajillo»
<i>Shrimps</i>	Camarones	Aparece en un elevado número de ocasiones. Variante latinoamericana, en España se dice «gambas»
<i>Drinks</i>	Tragos	Variante latinoamericana, en ES «bebidas»
<i>Husband / wife</i>	Esposo / esposa	Aunque la palabra no es incorrecta, en español de España se suele utilizar más la palabra «marido» y la palabra «mujer». Este error se da en todos los casos en que en inglés aparecen las palabras <i>husband</i> y <i>wife</i> .
<i>Waitress</i>	Mesera	Variante latinoamericana, en ES «camarera»

### 3.2.2. Análisis cualitativo de los resultados de la TA en el corpus de reseñas de restaurantes

Además, de analizar las diferentes categorías de errores y el número de estos en el corpus, el análisis ha permitido evaluar los niveles de calidad de la TA al español de las reseñas, clasificándolas en los niveles «nativa», «correcta», «aceptable» o «inaceptable». En la

Ilustración 21, se han proporcionado los datos relativos al número y a los porcentajes correspondientes a los diferentes grados de aptitud de las reseñas de restaurantes:

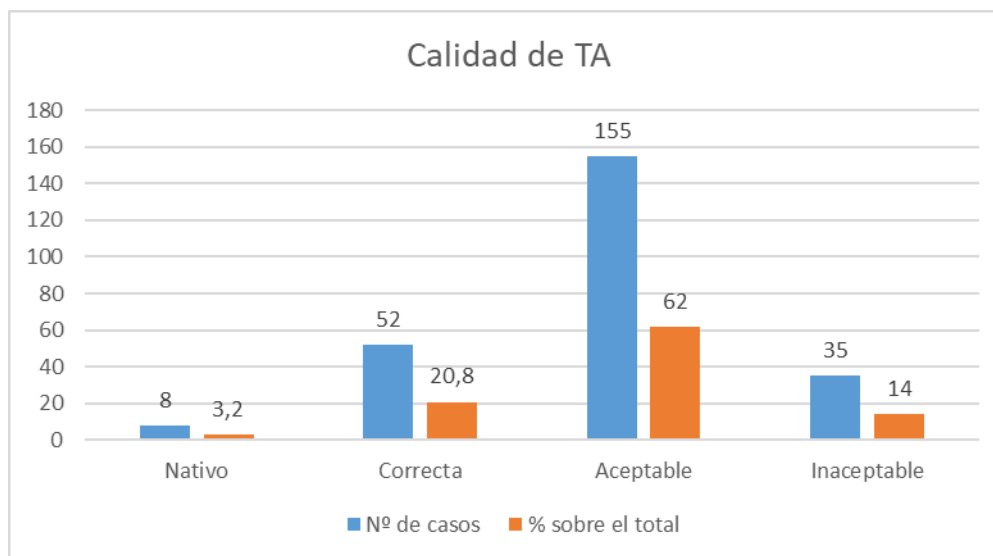


Ilustración 21. Niveles de calidad de las reseñas sobre restaurantes traducidas con TA.

El análisis sobre los grados de calidad de TA de las reseñas (Ilustración 21), ha mostrado que el nivel de calidad de más de la mitad de las reseñas (62 %) es aceptable. Algo especialmente relevante puesto que, aunque dichas reseñas no estén escritas originalmente en español y, por ende, no sean siempre correctas a nivel sintáctico, morfológico o gramatical, estas sí que pueden servir para que el lector pueda crearse una idea general del producto o servicio ofertado. Un caso similar es el de las reseñas con nivel de calidad correcta. Estas son las segundas más numerosas, con un porcentaje del 20,8 % con respecto al total. En estos casos, aunque la reseña no sea completamente correcta porque contiene errores lingüísticos, esta se puede entender.

Con respecto al porcentaje de reseñas de calidad inaceptable, este es del 14 % con respecto al total, porcentaje que, aunque no parezca muy elevado, es lo suficientemente significativo para que la reseña pueda perjudicar la imagen del restaurante. Si el usuario no entiende lo que está escrito en la reseña, este no podrá hacerse una idea de lo que va a adquirir o del lugar que va a visitar, por lo que podría decidirse por otras opciones en vez de por la que se oferta. Por este motivo, poseer el contenido adecuadamente puede llegar a ser decisivo. Por último, el porcentaje de reseñas con un nivel de calidad nativo es de tan solo el 3,2 %, porcentaje que podría ser mucho mayor si dichas reseñas se poseeraran y localizaran correctamente.

Adicionalmente, se han examinado los aspectos que afectan al nivel de calidad de la TA, según los errores de tipo cultural que esta ha cometido y que afectan sobre todo a alimentos regionales, en el caso de los restaurantes, y a lugares turísticos, en el caso de los hoteles. Es especialmente relevante tener este aspecto en cuenta cuando se traza una estrategia de



localización y para estudiar cómo mejorar la imagen de marca gracias a la posesión de las reseñas.

### 3.2.2.1. Problemas lingüísticos de localización que afectan a los niveles de calidad de la TA. *Corpus restaurantes*

A continuación, la Tabla 19, muestra los resultados extraídos del análisis del corpus de reseñas de restaurantes con errores de localización de tipo lingüístico:

*Tabla 19. Clasificación de la calidad de la TA en función del número y porcentaje de errores lingüísticos. Corpus restaurantes.*

<b>CALIDAD TA</b>	<b>Nº errores LINGÜÍSTICOS</b>
<b>Nativo</b>	1
<b>Correcta</b>	20
<b>Aceptable</b>	73
<b>Inaceptable</b>	14

En este caso, los resultados sobre el corpus, resumidos en la Tabla 19, muestran que la gran mayoría de las traducciones de las reseñas de restaurantes con errores lingüísticos de localización cuentan con un nivel de calidad aceptable (73 sobre el total), en segundo lugar, se encuentran las reseñas con un nivel de calidad correcto (un total de 20), a continuación, se encuentran las reseñas con nivel de calidad inaceptable (un total de 14) y, por último, las reseñas con un nivel nativo (solo 1). Resulta interesante que la gran mayoría de las reseñas sean aceptables, ya que, como se ha mencionado anteriormente, aunque no sean nativas, sirven para hacerse una idea del contenido y no impiden la comprensión. Además, el hecho de que solo haya una reseña con calidad nativa y que, por el contrario, haya un total de 14 reseñas inaceptables es muy significativo, ya que podría ser un dato muy descriptivo sobre la situación real de la calidad de la TAN.

La Ilustración 22, a continuación, muestra los porcentajes correspondientes a las reseñas del corpus de restaurantes con errores lingüísticos que afectan a los niveles de calidad de TA:

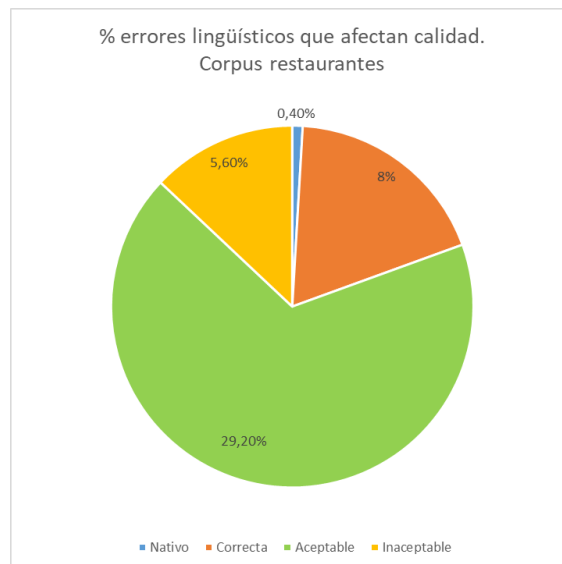


Ilustración 22. Porcentaje de errores lingüísticos que afectan a la calidad de la TA. Corpus restaurantes.

Como se puede apreciar en la Ilustración 22, el 29,2 % de los resultados de TA de las reseñas de restaurantes tienen un nivel de calidad aceptable. El total de reseñas cuyas traducciones automáticas tienen un nivel de calidad correcto representa el 8 %. Por otro lado, las traducciones con un nivel de calidad inaceptable representan el 5,6 % del total y, por último, la única reseña con errores lingüísticos pero que aun así cuenta con un nivel de calidad nativo representa el 0,4 % del total. Algo que resulta curioso es que haya más reseñas con un nivel de calidad aceptable, aunque tengan errores lingüísticos y que, además, este porcentaje sea más elevado que las reseñas con un nivel de calidad inaceptable. Esto puede deberse al hecho de que hay errores lingüísticos más graves que otros y algunos de ellos pueden no llegar a afectar en gran manera a la traducción automática resultante.

En la categoría de errores ilustrados en la Ilustración 22, han aparecido ciertos errores lingüísticos recurrentes, que afectan al nivel de calidad de la TA y que se han enumerado en la Tabla 20, a continuación:

Tabla 20. Diferentes categorías de errores lingüísticos.

Original	TA	Comentario
<b>Vº «tener» con comida</b>		
<i>We had paella and wine</i>	Teníamos paella y vino	En todos estos casos se realiza un uso incorrecto en ES de este verbo. Serían más comunes verbos como «tomar» o «comer»
<i>Meals I've had</i>	Comidas que he tenido	
<i>I had coffee</i>	Tuve un café	
<i>We had one with duck</i>	Tuvimos una con pato	
<i>We had carpaccio</i>	Tuvimos carpaccio	
<i>We had a great lunch</i>	Tuvimos un gran almuerzo	

<b>Uso equivocado de los verbos «ser/estar»</b>		
<i>Paella was amazing</i>	La paella fue increíble	En todos estos casos se confunde el uso de estos verbos, ya que, para este contexto en ES se utilizaría el verbo «estar»
<i>Paella was excellent and wine too</i>	Paella era excelente y vino también	
<b>Falta de concordancia de adjetivos</b>		
<i>Paella [...], which was delicious</i>	La <u>paella</u> [...] que estaba <u>delicioso</u>	«Paella» es un sustantivo femenino y los adjetivos en estos ejemplos son masculinos. En estos casos se han traducido incorrectamente los adjetivos, lo que ha producido una falta de concordancia de género con el sujeto.
<i>Paella was excellent (although the waiter served a different one from we ordered)</i>	<u>Paella</u> era excelente (aunque un camarero sirvió <u>uno</u> diferente de lo que pedimos)	
<i>Paella was amazing! So flavourful and cooked to perfection.</i>	La <u>paella</u> fue increíble! Tan <u>sabroso y cocinado</u> a la perfección	
<b>Oraciones de relativo</b>		
<i>The second time we visit</i>	la segunda vez visitamos	Formación incorrecta de la oración de relativo. Debería ser «la segunda vez <u>que</u> visitamos».
<i>Was recommended to us after we took a boat trip</i>	Nos lo recomendaron <u>después de que hicimos</u> un paseo en bote	Formación incorrecta e innecesaria de la oración de relativo. Debería ser «nos lo recomendaron <u>después de haber</u> dado un paseo en barca».
<b>Cambios de sentido</b>		
<i>Filled with rich shrimp [...] and cuttlefish</i>	<u>Llena de</u> ricos camarones [...] y <u>sepias</u>	Traducción incorrecta que cambia el significado y da a entender que el plato tenía muchas sepias, cuando en realidad lo que implica es que lleva «muchos trozos de sepia», algo que solo se sabe si se es nativo de la zona y se conoce el plato.
<i>The service was friendly (and did not hover)</i>	El <u>servicio</u> fue amable (y no <u>flotaba</u> )	Traducción incorrecta que cambia el significado y deja sin sentido la oración que se refería a que el servicio no era agobiante.
<i>We were lucky to be received without reservation</i>	Tuvimos la suerte de ser recibidos <u>sin reservas</u>	Traducción incorrecta que cambia el significado. El texto se refería a «sin tener reserva».
<b>Otros ejemplos</b>		
<i>The restaurant was a gem</i>	El restaurante <u>era una joya</u>	Traducción literal de la expresión en inglés.
<i>Seafood paella</i>	Paella marisco	Paella <u>DE</u> marisco
<i>Try paella</i>	<u>Experimentar</u> la paella	<u>Probar</u> la paella

<i>Very good paella sitting outside the beautiful terrace.</i>	Muy buena <u>paella sentada</u> afuera de la hermosa terraza.	Traducción incorrecta. El original expresa que la persona se comió la paella estando sentada en una terraza.
<i>Was a few steps ahead</i>	está varios pasos por delante	Traducción literal de la expresión en inglés.
<i>Table clothes</i>	Ropa de mesa	Traducción literal incorrecta. Traducción correcta sería «mantelería».

### 3.2.2.2. Problemas culturales de localización que afectan a los niveles de calidad de la TA. Corpus restaurantes

La Tabla 21 muestra que la mayoría de los resultados de TA del corpus de restaurantes son aceptables (28 sobre el total de 250). Sin embargo, tan solo 10 de las traducciones son correctas, 6 son consideradas nativas y solo 6 son inaceptables.

Tabla 21. Clasificación de la calidad de la TA en función del número y porcentaje de errores culturales. Corpus restaurantes.

CALIDAD TA	Nº errores CULTURAL
<b>Nativo</b>	6
<b>Correcta</b>	10
<b>Aceptable</b>	28
<b>Inaceptable</b>	6

Resulta interesante que el número de traducciones con un nivel aceptable sea mayor que el del resto de niveles de calidad, ya que, aunque las reseñas traducidas al español no se hayan escrito originalmente en este idioma, son útiles para que el lector pueda formarse una idea general de la experiencia en el restaurante y para que pueda tomar la decisión de ir a ese restaurante o no. Esto es claramente visible en la Ilustración 23, a continuación, que contiene los porcentajes mencionados anteriormente con respecto al corpus de reseñas de restaurantes:

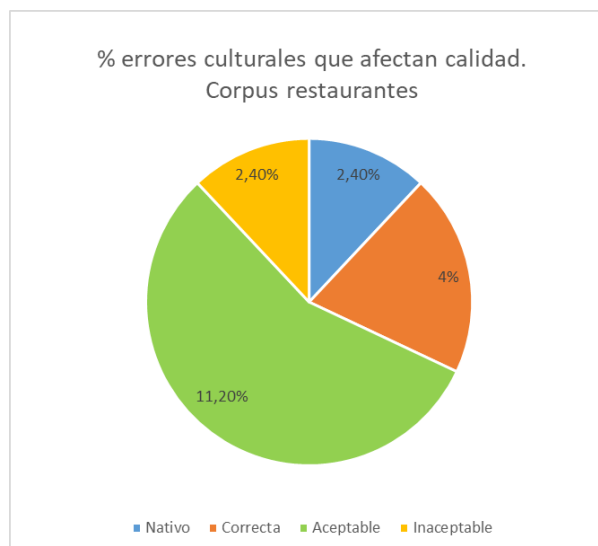


Ilustración 23. Porcentaje de errores culturales que afectan a la calidad de la TA. Corpus restaurantes.

Como se aprecia en la Ilustración 23, un 11,2 % de las traducciones de las reseñas se consideran aceptables, solo el 4 % son correctas, otro 4 % se consideran nativas y solo el 2,4 % tienen un nivel de calidad inaceptable.

La mayoría de los ejemplos identificados de errores de tipo cultural en el caso de los restaurantes han sido palabras, sobre todo relacionadas con la comida, que la TA ha traducido de maneras distintas, pese a que en inglés era el mismo concepto. Es más, en ocasiones, esto ocurre incluso cuando esta misma palabra aparece repetidas veces en una misma reseña. Algunos ejemplos se han recogido en la Tabla 22, a continuación:

Tabla 22. Errores culturales. Ejemplos de una misma palabra en inglés traducida inconsistentemente. Corpus restaurantes.

Original	TA
<i>Shrimp</i>	langostinos, gambas o camarones
<i>Gambas al AJILLO</i>	gambas con ají, gambas al ajillo, gambas all'ajillo
<i>Squid</i>	Calamari
<i>Potatoes</i>	Papas, patatas

La TA presenta además problemas para identificar los platos tradicionales de la región, y esto da como resultado traducciones como las que aparecen en la Tabla 23:

Tabla 23. Errores culturales. Ejemplos de errores con los nombres de platos regionales. Corpus restaurantes.

Original	TA	Comentario
----------	----	------------

<i>Paella with carabineers</i>	Paella con carabiners	Traducción correcta «paella de marisco».
<i>Ali y oli</i>	Ali y oli	Traducción correcta «alioli» o «all i oli», dependiendo de si se dice utiliza la variante adaptada al español o si se utiliza en español la palabra en valenciano. Ambas opciones son correctas.
<i>Initial dishes</i>	Platos iniciales	Traducción correcta «entrantes».
<i>Dirty squid</i>	Calamar sucio	En realidad se trata de «sepia sucia» o «sepia de playa», muy típica de la zona.
<i>A soupy paella</i>	Paella espesa	Plato correctamente denominado «arroz meloso».

Este problema puede darse en parte porque muchos de los platos tienen nombres en valenciano, como es el caso de «Pa amb Oli», que en realidad se refiere a «pà amb all i oli».

Estos errores no solo aparecen con respecto a la comida, sino que también con relación a lugares turísticos, como L'Albufera, u otras denominaciones o conceptos regionales. De esta manera, la TA produce resultados como «la laguna» o «el lago»; el almuerzo (en Valencia se denomina «la comida» a lo que se toma a mediodía, no «almuerzo», ya que el almuerzo es otra comida que se realiza en la región a media mañana). Otro claro ejemplo de este tipo de errores de localización en el caso del español de España es el uso por parte de la TA de palabras utilizadas en el español de Latinoamérica como «mesera» (que en España se denomina «camarera») o «trago» (que en España se diría «bebida»).

### ***3.2.2.3. Causas frecuentes que afectan al nivel de calidad en el corpus de reseñas de restaurantes***

La calidad de la TA también se ve afectada por otros factores como los errores que ya contenían las reseñas originales, los errores que comete el motor de TA a causa de la calidad de este o los problemas que se han denominado de duplicidad o explicación, es decir, que en las reseñas traducidas con TA se han repiten conceptos que aparecían entre paréntesis en las reseñas originales porque en inglés necesitaban una explicación para entenderse. Sin embargo, estas explicaciones, son redundantes e innecesarias en español, ya que se refieren a conceptos

reconocidos en el *locale* español. Todos estos factores se han analizado y los resultados se presentan en la Tabla 24, a continuación:

Tabla 24. Número problemas que afectan al nivel de calidad de la TA. Corpus restaurantes.

CAUSAS QUE AFECTAN A LA CALIDAD DE LA TA	Nº errores
Problemas explicación / duplicidad	217
Problemas calidad motor TA	172
Problemas error en el original	41

En la Tabla 24 anterior se aprecian los tipos de problemas que afectan a la calidad de la TA de las reseñas del corpus de restaurantes. En primer lugar, los problemas de explicación y duplicidad han sido los más numerosos con un total de 217 casos. En segundo lugar, han surgido problemas a causa de la calidad del motor de TA, en total se han dado 172 casos. Por último, también se han contemplado los casos de problemáticas a causa de errores ya presentes en las reseñas originales, con un total de 41 casos. Si se observan los porcentajes relativos a estos datos, se puede apreciar lo significativos que pueden ser estos tipos de errores (Ilustración 24):

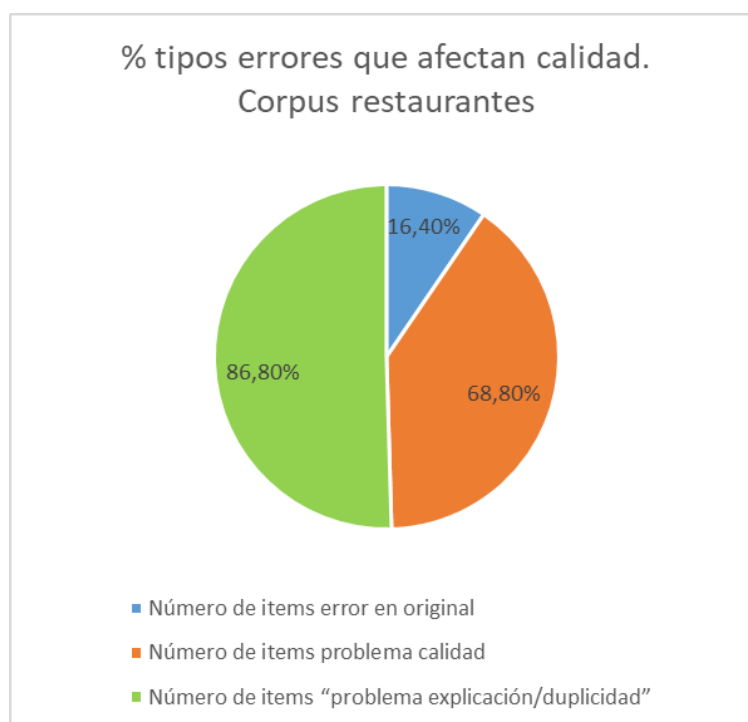


Ilustración 24. Porcentaje factores que afectan al nivel de calidad de la TA. Corpus restaurantes.

En la Ilustración 24, se han destacado diversos aspectos que han afectado a la calidad de las traducciones de manera negativa, por lo que la naturalidad de estas se ha visto reducida y, en consecuencia, se ha perjudicado la credibilidad y la percepción de la reseña. Por un lado, los errores más representativos, dado el porcentaje de casos (85,80 %), han sido los problemas de

duplicidad o explicación. Este tipo de errores surge porque la TA repite el concepto o explicación que en el original en inglés estaba entre paréntesis porque requería una aclaración por no ser parte del vocabulario o de la cultura angloparlante. La Tabla 25 siguiente presenta algunos ejemplos de duplicidad o explicación encontrados en el corpus de restaurantes:

Tabla 25. Ejemplos de errores de duplicidad o explicación encontrados. Corpus de restaurantes.

Original	TA
Seafood Paella/Paella Marisco	Paella Marisco / Paella Marisco
Fabulous clochinas (local mussels)	Las fabulosas clochinas (mejillones locales),
best arroz de marisco (paella)	mejores arroz de marisco (paella)
chipirrones (baby squid)	chipirrones (chipirones)
carabineros (prawns)	carabineros (langostinos)
fideua (which is pasta)	fideua (que es pasta)
arroz al horno [...] (rice in the oven)	arroz al horno [...] (arroz al horno),
clóchinas (mussels)	clóchinas (mejillones)
black pudding (morcilla)	morcilla (morcilla)
revuelto (scrambled eggs)	revuelto (huevos revueltos)

En la Tabla 25, se observa que estas duplicidades o explicaciones pueden dar lugar a diferentes errores. Por un lado, en ocasiones, la misma palabra se repite dos veces en los textos traducidos con TA porque en la reseña original se mencionaba un plato regional valenciano en inglés y luego, entre paréntesis, se introducía el nombre original en español. La TA no sabe distinguir que se trata de un mismo concepto, lo que resulta en traducciones como «*Seafood Paella / Paella Marisco* → Paella Marisco / Paella Marisco». Sin embargo, el error más recurrente dentro de esta categoría es el de la explicación de un concepto o, en este caso, alimento: «carabineros (langostinos)» o «clóchinas (mejillones)».

En estos casos, la TA introduce entre paréntesis un concepto equivocado que no corresponde a lo que se ha querido expresar en el original, ya que no es lo mismo un carabinero que un langostino, ni las clóchinas que los mejillones. En este caso, puesto que para un hablante de español estas explicaciones son innecesarias, lo correcto hubiera sido eliminar los paréntesis. Por último, dentro de esta categoría, se encuentra el caso de los adjetivos que se repiten en español cuando en inglés se utilizan adjetivos sinónimos. Por ejemplo: *very gentle, very kind behavior* => un comportamiento muy amable, muy amable; *delightful, delicious* => delicioso, delicioso o *yum yum yummy food* => deliciosa comida deliciosa.

En segundo lugar, se ha detectado la problemática relacionada con la calidad del motor de TA utilizado. A continuación, en la Tabla 26, se enumeran algunos ejemplos de estos errores:



Tabla 26. Ejemplos de errores causados por la calidad del motor de TA. Corpus de restaurantes.

Original	TA
<i>Paella was excellent [...] and rose wine too</i>	Paella era excelente [...] y vino rosado también
<i>Sporting clients</i>	Cientes deportivos
<i>Off the chain good</i>	Fuera de la cadena bueno
<i>We took the 25 minute drive</i>	Tomamos el viaje de 25 minutos
<i>Overall reasonable value for money</i>	Valor razonable en general para el dinero
<i>I am gluten and dairy free</i>	No tengo ni gluten ni lácteos
<i>Visited here</i>	Visité aquí
<i>Best fideuá around</i>	Mejor fideuá alrededor
<i>Thus spake a vegetarian</i>	Por lo que habló un vegetariano
<i>Contributed to my enjoyment</i>	Contribuyó a mi disfrute
<i>Optional wine flight</i>	Vuelo opcional

Como muestra la Tabla 26, estos errores suelen producirse por la traducción literal de algunas expresiones idiomáticas del inglés como *off the chain good*, *as good as it gets* o *out of this world*. De nuevo, se observa una utilización errónea de los verbos «ser» y «estar» cuando en inglés se utiliza el verbo *to be* en relación con la calidad o a alguna cualidad de la comida, así como del verbo «tener» (*to have*) cuando el usuario relata los platos que probó. A continuación, se han proporcionado dos ejemplos encontrados en el corpus de restaurantes:

- **Verbo «tener»:** «tuve un café», «teníamos paella y vino» o «tuvimos tapas».
- **Verbos «ser» y «estar»:** «la paella fue increíble», «las papas eran» o «paella era excelente».

Por último, la Tabla 27, a continuación, refleja algunos ejemplos de errores que se han producido en la TA al español cuando en el original ya había errores. Estos problemas pueden ser desde palabras mal escritas que no se han corregido en la traducción, hasta oraciones que, por un ligero error, cambian totalmente el significado.

Tabla 27. Ejemplos de errores causados por errores en la reseña original en inglés. Corpus de restaurantes.

Original	TA	Comentario
<i>Epecially pellas</i>	Especialmente pellas	Se refiere a «paellas»
<i>Our waitress but he was nice</i>	Nuestra camarera, pero fue encantador	Falta de concordancia entre sujeto y adjetivo

<i>Paella e marisco</i>	Paella e marisco	Se refiere a «paella de marisco»
<i>Socarrat</i>	Socarrat	Se refiere a «socarrat», al haberse escrito incorrectamente en el original, la TA lo ha dejado como estaba en vez de utilizar la escritura correcta de la palabra
<i>6: 25 euros</i>	6: 25 euros	Debería utilizar una coma «6,25 euros»
<i>Grain of reice keeps</i>	Cada grano de conserva	Se escribe la palabra <i>rice</i> incorrectamente y la TA lo omite, por lo que al final ofrece una traducción completamente errónea
<i>Sarters</i>	Sarters	Se refiere a <i>starters</i> que en español corresponde a «entrantes»
<i>The shorts (shots)</i>	Los cortos (chupitos)	Se refiere a <i>shots</i> que en español corresponde a «chupitos»
<i>The croquet were Excellent</i>	El croquet estuvo excelente	Se refiere a «croquetas»
<i>The octupos</i>	Los octupos	Se refiere a <i>octopus</i> que en español corresponde a «pulpo»
<i>Baby sqid</i>	Sqid bebé	Se refiere a <i>baby squid</i> que en español corresponde a «chipirón»

<i>Cheese and meat plateau</i>	Meseta de queso y carne	Se refiere a <i>cheese and meat plate</i> que en español corresponde a «tabla de queso y embutidos»
<i>Must try the dirty squid and cuttlefish Best restaurant</i>	Debe probar el mejor restaurante sucio de calamar y sepia que	El autor no escribe el punto y seguido en la reseña original, y la TA ofrece un resultado erróneo. «Debe probar la sepia sucia y los calamares. El mejor restaurante»

Un dato interesante es que el motor de TA ha corregido de manera arbitraria algunos de los errores que aparecían en las reseñas originales en inglés, es decir, en algunos casos ha corregido estos errores, pero en otros no. Además, se han detectado casos en que, en la reseña traducida, se omiten palabras del original que estaban escritas incorrectamente. Se han encontrado los ejemplos siguientes (Tabla 28):

Tabla 28. Ejemplos de errores por palabras mal escritas. Corpus de restaurantes.

Original	TA	Comentario
<i>grain of <u>reice</u> keeps</i>	cada grano de conserva	Omite «de arroz»
<i>a <u>reaalu</u> good paella</i>	una paella buena	El <i>really</i> no se ha traducido y se pierde, por tanto, este matiz en la traducción)
<i>potatoesEel</i>	Laanguila	Las patatas se han omitido y se traduce el artículo y el sustantivo sin espacio entre sí

### 3.3. Análisis cuantitativo y cualitativo del corpus de reseñas sobre hoteles traducidas con TA

Al igual que en el corpus de reseñas de restaurantes, primero, se ha llevado a cabo un análisis de datos generales sobre localización, en que se han estudiado los errores culturales y los lingüísticos. Tras realizar el análisis cuantitativo de dichos errores de localización, se ha llevado a cabo el análisis cualitativo para determinar en qué nivel afectan estos problemas de localización a la calidad de los resultados de traducción.

### 3.3.1. Análisis cuantitativo de aspectos de localización en el corpus de reseñas de hoteles

La Tabla 29, a continuación, contiene los datos de los mismos errores mencionados en el apartado del análisis cuantitativo de las reseñas de restaurantes, pero en este caso se hace referencia al corpus de reseñas de hoteles:

Tabla 29. Datos sobre número de reseñas sobre hoteles y los principales tipos de errores de localización.

Total de reseñas	250
Nº reseñas con errores culturales	101
Nº reseñas con errores lingüísticos	209
<b>Total de errores de localización</b>	<b>310</b>

En el corpus de reseñas de hoteles ha surgido un total de 310 errores de localización, de los cuales, 101 han sido errores de localización culturales y, 209 han sido errores de localización de tipo lingüístico. La Ilustración 25 muestra los porcentajes representativos de estos datos:

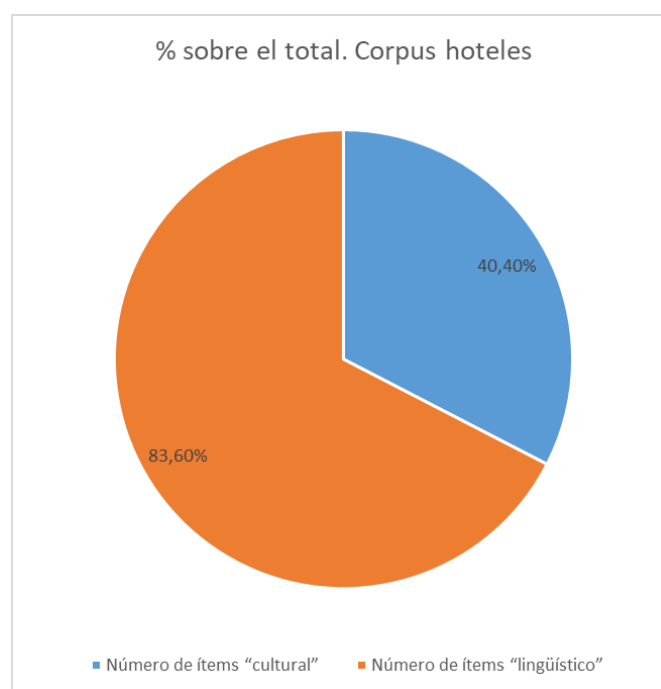


Ilustración 25. Datos sobre el porcentaje de reseñas sobre hoteles y los principales tipos de errores de localización.

En cuanto a los errores de tipo cultural detectados (40,40 %), se aprecia que estos tienen su origen en que la TA no traduce correctamente los lugares históricos o turísticos de Valencia, así como los nombres de los barrios, las calles u otras denominaciones. En la Tabla 30, a continuación, se muestran algunos de estos ejemplos:

Tabla 30. Ejemplos de lugares localizados incorrectamente.

Original	TA	Comentario
<i>The Nord train</i>	Del tren Nord	Traducción correcta «la Estación del Norte».
<i>Maritin Serreria</i>	Maritin Serreria	Error en el original que causa una traducción incorrecta. La denominación correcta es «Marítim Serrería».
<i>The green belt park</i>	Parque del cinturón verde	Denominación correcta «parque del cauce del río Turia».

En su mayoría suelen ser errores que se producen porque la TA emplea palabras de las variantes latinoamericanas del español (piso laminado => suelo laminado; computadora portátil => ordenador portátil; huelen divino => huelen divinamente/genial; desempacar => deshacer la maleta; a una cuadra => a una manzana). Además, se han encontrado muchas expresiones no traducidas del inglés que en español no se entienden por falta de contexto o porque estas expresiones no tienen equivalente directo en esta lengua: *without fake American smiling* → «sin falsa sonrisa estadounidense»; *hop on hop off bus* → «autobús Hop on hop off»; *Buck's Fizz and lite English Breakfast* → «Buck's Fizz y lite English Breakfast».

En cuanto a los errores lingüísticos (83,60 %) que afectan a la localización, al igual que en el caso de los restaurantes, en nuestro corpus, la TA emplea incorrectamente el verbo «tener» en relación con la comida, pero, en esta ocasión, también en otros contextos, cuando en inglés aparece el verbo *to have*. Se han expuesto algunos ejemplos en la Tabla 31 siguiente:

Tabla 31. Traducción incorrecta del verbo *to have*. Corpus hoteles.

Original	TA	Comentario
<i>We <u>had</u> a break</i>	<u>Tuvimos</u> un descanso	<u>Tomamos</u> un descanso
<i>We <u>had</u> a breakfast</i>	<u>Tuvimos</u> un desayuno	<u>Desayunamos/Tomamos</u> un desayuno
<i>We <u>had</u> value for money</i>	<u>Teníamos</u> una buena relación calidad-precio	Lo que tiene buena calidad es el producto no las personas que adquieren el

producto. En este caso «El hotel ofrecía una buena relación calidad-precio».

<i>I <u>had</u> some complimentary champagne</i>	<u>Tuve</u> un poco de champán	<u>Tomé</u> un poco de champán / <u>Bebí</u> un poco de champán
<i>We <u>have</u> light breakfasts at home</i>	<u>Tenemos</u> desayunos ligeros en casa	<u>Desayunamos</u> ligero / <u>Tomamos</u> desayunos ligeros
<i>We <u>had</u> a hire car</i>	<u>Teníamos</u> un coche de alquiler	La expresión no es incorrecta en sí, no obstante, suena mucho más natural « <u>Alquilamos</u> un coche»
<i>We <u>had</u> free parking</i>	<u>Tuvimos</u> aparcamiento gratuito	<u>Había</u> servicio de aparcamiento gratuito
<i>They <u>have</u> weddings and business conferences</i>	<u>Tienen</u> bodas y conferencias de negocios	<u>Celebran</u> bodas y conferencias de negocios
<i>What we experienced there, we never <u>had</u> it somewhere else</i>	Lo que experimentamos allí, nunca lo <u>tuvimos</u> en otro lugar	Esto no nos pasó en ningún otro lugar
<i>To <u>have</u> valuable family time together</i>	<u>Tener</u> un valioso tiempo en familia juntos	<u>Pasamos</u> un valioso tiempo en familia

De la misma manera, la TA comete errores en el uso de los verbos «ser» y «estar» en relación con la comida, pero también en otros contextos, cuando en inglés aparece el verbo *to be*. La Tabla 32, a continuación, contiene algunos ejemplos encontrados en el corpus:

Tabla 32. Traducción incorrecta del verbo *to be*. Corpus hoteles.

Original	TA	Comentario
<i>The breakfast was standard, <u>wasn't</u> diverse enough in my opinion</i>	El desayuno era estándar, no <u>estaba</u> lo suficientemente diverso en mi opinión.	En este caso la TA incluso traduce bien el primer <i>was</i> e incorrectamente el segundo. «Era un desayuno estándar, que en mi opinión no era lo suficientemente variado»

<i>Breakfast <u>was</u> rich</i>	Desayuno es rico	En este caso comete dos errores: 1. Traduce el adjetivo incorrectamente 2. Conjuga el verbo en un tiempo incorrecto  Traducción correcta «El desayuno era abundante»
<i>Room <u>was</u> clean</i>	La habitación <u>era</u> limpia	Confunde el verbo «ser» y «estar». Traducción correcta «La habitación estaba limpia»
<i>The buffet breakfast <u>was</u> good value</i>	El desayuno buffet <u>era</u> una buena relación calidad-precio	Ha confundido el uso del verbo «ser» con el «tener». El desayuno buffet <u>tenía</u> una buena relación calidad-precio
<i>Breakfast <u>was</u> delicious</i>	El desayuno <u>era</u> delicioso	El desayuno <u>estaba</u> delicioso
<i>Paella <u>was</u> amazing</i>	La paella <u>era</u> increíble	La paella <u>estaba</u> increíble
<i>Not much of a view</i>	No es una gran vista	No <u>tiene/cuenta con</u> una gran vista
<i>Standard rooms which <u>were</u> a decent size</i>	Habitaciones estándar que <u>eran</u> de un tamaño decente	Ha confundido el uso del verbo «ser» con el «tener». «Habitaciones estándar que <u>tenían</u> un tamaño decente»
<i>Balcony which <u>was</u> of good size</i>	Balcón que <u>era</u> de buen tamaño	Ha confundido el uso del verbo «ser» con el «tener». «Balcón que <u>tenía</u> un buen tamaño»
<i>Room <u>was</u> of a nice size</i>	La habitación <u>era</u> de buen tamaño	Ha confundido el uso del verbo «ser» con el «tener». «La habitación <u>tenía</u> un buen tamaño»

---

Por último, la TA traduce de manera literal las expresiones idiomáticas provenientes del inglés, como, por ejemplo: *Good value for money* → buen valor para el dinero; *Makes the difference* → hace la diferencia; *is sensational value* → tiene un valor sensacional.

### 3.3.2. Análisis cualitativo de los resultados de la TA en el corpus de reseñas de hoteles

En el caso de los hoteles, representado en la Ilustración 26 siguiente, los resultados no distan mucho de los obtenidos en el análisis de las reseñas de restaurantes:

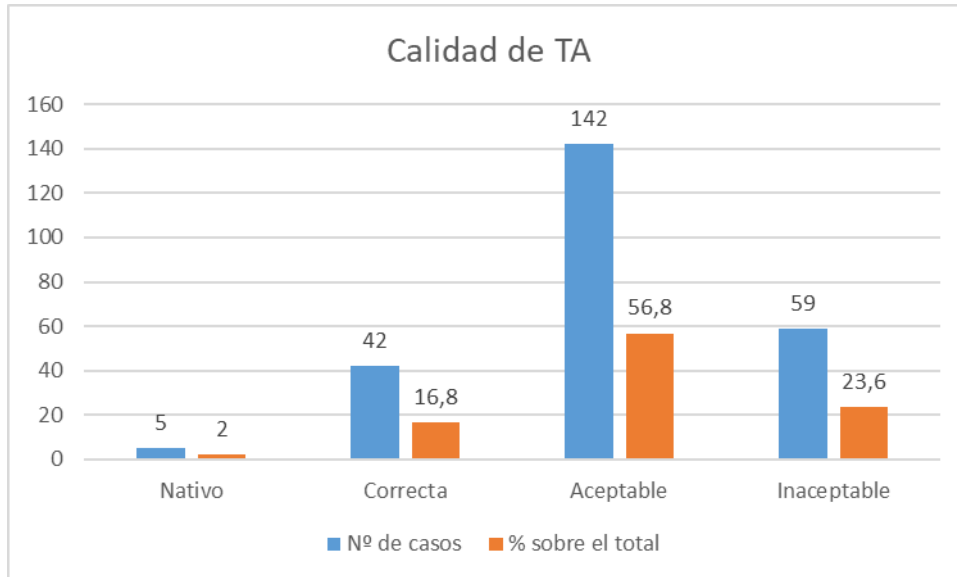


Ilustración 26. Niveles de calidad de las reseñas sobre hoteles traducidas con TA.

Al igual que en el caso de los restaurantes, la mayoría de las reseñas de hoteles presentan un nivel de calidad aceptable, un 56,8 % del total, aunque en este caso el porcentaje es algo más bajo, no es una diferencia significativa. La gran diferencia, sin duda se encuentra, en que, en este caso, el porcentaje de reseñas con un nivel inaceptable es mayor que en el caso de los restaurantes, pasando de un 14 %, en el caso de las reseñas de restaurantes, a un 23,6 %. Esto ha repercutido sobre todo en el porcentaje total de reseñas con un nivel de calidad correcto (16,8 %), que en esta ocasión, es menor que el porcentaje en el caso de las reseñas de restaurantes. Por último, también se ha visto reducido el porcentaje de reseñas con un nivel de calidad nativo, que en este caso representa solo un 2 % del total.

Adicionalmente, se han examinado los aspectos que afectan al nivel de calidad de la TA, según los errores de tipo cultural que esta ha cometido y que afectan sobre todo a nombres de alimentos regionales, en el caso de los restaurantes, y a lugares turísticos, en el caso de los hoteles. Es especialmente relevante tener en cuenta este aspecto cuando se elabora una estrategia de localización.



### 3.3.2.1. Problemas lingüísticos de localización que afectan a los niveles de calidad de la TA. Corpus hoteles

A continuación, la Tabla 33, muestra los resultados del análisis del corpus de reseñas de hoteles con errores de localización de tipo lingüístico:

Tabla 33. Clasificación de la calidad de la TA en función del número y porcentaje de errores lingüísticos. Corpus hoteles.

<b>CALIDAD TA</b>	<b>Nº errores LINGÜÍSTICOS</b>
<b>Nativo</b>	1
<b>Correcta</b>	35
<b>Aceptable</b>	115
<b>Inaceptable</b>	56

En estos resultados del análisis del corpus de reseñas de hoteles con errores de localización de tipo lingüístico, vuelve a destacar el número de reseñas con nivel de calidad nativo que, al igual que en el caso de las reseñas de restaurantes, es de solo 1. De igual forma, el número de reseñas cuya traducción tiene un nivel de calidad aceptable es muy superior al resto con un total de 115 reseñas. A este dato, le siguen las reseñas con traducciones cuya calidad es inaceptable, con 56 reseñas, y, por último, las reseñas con traducciones automáticas de nivel correcto, que son solo 35.

Tanto en el corpus de hoteles como en el de restaurantes, la mayor parte de reseñas con errores de localización lingüísticos tienen un nivel de calidad aceptable y las menos numerosas son las nativas, que en ambos casos cuentan con solo un resultado. No obstante, la gran diferencia se encuentra en que mientras que en el caso de los restaurantes las reseñas con traducciones de calidad correcta eran las segundas más numerosas, en el caso de los hoteles, el segundo lugar es para las reseñas con traducciones de calidad inaceptable.

La Ilustración 27 expone el porcentaje de errores de localización lingüísticos correspondientes a los niveles de calidad de la TA en el caso del corpus de reseñas de hoteles:



Ilustración 27. Porcentaje de errores lingüísticos que afectan a la calidad de la TA. Corpus hoteles.

Como se puede apreciar en la Ilustración 27, el 46 % de las TA de las reseñas de hoteles tienen un nivel de calidad aceptable. El total de reseñas cuyas traducciones automáticas tienen un nivel de calidad correcto representa el 14 %. Por otro lado, las traducciones con un nivel de calidad inaceptable representan el 22,40 % del total y, por último, las reseñas con nivel de calidad nativo representan el 0,4 % del total. En este caso, como se puede apreciar, el porcentaje de reseñas con nivel inaceptable es mucho mayor que en el caso de los restaurantes.

Como se muestra en la Ilustración 27, han surgido ciertos errores lingüísticos recurrentes, que afectan al nivel de calidad de la TA y que se han enumerado a continuación (ver Tabla 34):

Tabla 34. Ejemplos de errores de localización lingüísticos. Corpus de hoteles.

Original	TA	Comentario
<b>Vº «tener» con comida</b>		
<i>We <u>had</u> breakfast</i>	<u>Teníamos</u> un desayuno.	En todos estos casos se realiza un uso incorrecto en ES de este verbo. «Tomamos un desayuno»
<i>I <u>had</u> some complimentary champagne</i>	<u>Tuve</u> un poco de champán.	«Tomé un poco de champán»
<i>We <u>had</u> a break.</i>	<u>Tuvimos</u> un descanso.	«Descansamos»
<i>What we experienced there, we never <u>had</u> it somewhere else.</i>	Lo que experimentamos allí, nunca lo <u>tuvimos</u> en otro lugar.	«Esto no nos ha pasado en ningún otro lugar»
<i>Able to have a massage</i>	Capaz de <u>tener</u> un masaje.	«Capaz de darme un masaje»
<b>Uso equivocado de los verbos «ser/estar»</b>		

<i>The breakfast was standard, <u>wasn't</u> diverse enough in my opinion</i>	El desayuno era estándar, no <u>estaba</u> lo suficientemente diverso en mi opinión.	El desayuno [...] no <u>era</u> lo suficientemente variado
<i>Breakfast <u>was</u> delicious ...you will be disappointed. I definitely <u>am</u>.</i>	Desayuno <u>era</u> delicioso ...se sentirá decepcionado. Definitivamente lo <u>soy</u> .	El desayuno <u>estaba</u> delicioso ...se sentirá decepcionado. Definitivamente lo estoy.
<i>The room <u>was</u> pristine</i>	La habitación <u>era</u> impecable	La habitación <u>estaba</u> impecable

#### Falta de concordancia de adjetivos y verbos

<i>Black out curtain - was perfect</i>	<u>Cortina</u> opaca, era perfecto	<u>Cortina</u> opaca, era perfecta
<i>Chemical cleaning fluids and spray - quite revolting</i>	<u>Líquidos</u> de limpieza químicos y <u>aerosoles</u> , bastante <u>repugnante</u>	<u>Líquidos</u> de limpieza químicos y <u>aerosoles</u> , bastante repugnantes
<i>However i felt this one was smaller than usual and was dated</i>	Estoy acostumbrado a <u>duchas</u> pequeñas, sin embargo, sentí que <u>este</u> era más <u>pequeño</u> y estaba <u>anticuado</u>	Estoy acostumbrado a duchas pequeñas, sin embargo, sentí que <u>esta</u> era más <u>pequeña</u> y estaba <u>anticuada</u>
<i>As a first time visitors to valencia, we wanted to stay</i>	Como <u>visitante</u> por primera vez a valencia, <u>queríamos quedarnos</u>	Como <u>visitantes</u> por primera vez a valencia, <u>queríamos quedarnos</u>
<i>Facilities were stunning and beautifully kept.</i>	Las <u>instalaciones</u> eran impresionantes y muy bien <u>cuidados</u>	Las <u>instalaciones</u> eran impresionantes y estaban muy bien cuidadas

#### Oraciones de relativo

<i>Every time I have received an excellent reception with the most amazing rooms</i>	Cada vez <u>que</u> he recibido una excelente recepción con las habitaciones más increíbles	Introduce un «que» que en este caso sobra. «Cada vez he recibido una excelente recepción con las habitaciones más increíbles»
<i>Epecially smiley blond lady receptionist which I sadly didn't take name of</i>	Especialmente sonriente recepcionista rubia <u>que</u> lamentablemente no tomé nombre	En este caso haría falta reformular la oración para que sonara natural «Especialmente la sonriente recepcionista rubia, <u>cuyo</u> nombre lamentablemente no anoté»
<i>The only thing that was missing was toilette brush, that also after I requested it, wasn't brought</i>	Lo único que faltaba era el cepillo de baño, que también después <u>de</u> <u>que</u> lo solicité, no fue traído	En este caso haría falta reformular la oración para que sonara natural ya que en este caso la estructura de la pasiva dificulta la formación «Lo único que faltaba era la escobilla, <u>que aunque la solicité</u> , no la trajeron»

<i>it's a room that I cannot wait to leave the second I step into it</i>	es una habitación <u>que</u> no puedo esperar para salir en el momento en que entro	«es una habitación <u>de la que</u> no pude esperar a salir desde el momento en que entré» o «es una habitación <u>de donde</u> no pude esperar a salir desde el momento en que entré»
<i>The first time we booked a suite, which I would highly recommend.</i>	La primera vez que reservamos una suite, <u>que</u> recomiendo encarecidamente	«La primera vez que reservamos una suite, <u>algo que</u> recomiendo encarecidamente»

---

**Cambios de sentido**

---

<i>Mixed review...</i>	Mezcla revisada	Traducción incorrecta que cambia el significado. En este caso sería necesario reformular y realizar un ejercicio casi de transcreación, ya que se trata de un título que da a entender que el autor de la reseña tiene diversas opiniones al respecto. Por ejemplo, podría traducirse «Reseña con sentimientos contradictorios».
<i>Only there for a meeting but staff so kind and efficient must try t stay a night!</i>	¡Solo para una reunión, pero el personal tan amable y eficiente debe tratar de pasar una noche!	La TA atribuye la acción de <i>must try to stay for a night</i> al personal lo que es incorrecto.
<i>Some of the hotel is a bit dated</i>	Algunos de los hoteles están un poco anticuados	TA errónea se refiere a que parte del hotel estaba anticuado no a que varios hoteles estaban anticuados
<i>the hotel itself was very busy on some days due to a wedding and what i think was a corporate event.</i>	el hotel en sí estaba muy ocupado algunos días debido a una boda y lo que hice. Creo que fue un evento corporativo	TA completamente errónea «el hotel en sí estaba muy ocupado algunos días debido a una boda y lo que creo que fue un evento corporativo»
<i>The rooms were extravagant and very comfortable</i>	Las habitaciones eran <u>extravagantes</u> y muy cómodas	Aunque la traducción no es incorrecta, la palabra «extravagante», puede llegar a tener connotaciones negativas en español, por lo que habría que sustituirla por otra que exprese algo más positivo.

También se han encontrado expresiones propias del inglés, como frases hechas o jerga, que se han traducido de manera literal o incorrecta, lo que da como resultado expresiones poco naturales en lengua meta. Algunos ejemplos se exponen en la Tabla 35, a continuación:

Tabla 35. Traducción incorrecta de frases hechas o expresiones en inglés. Corpus de hoteles.

Original	TA	Comentario
<i>Made this city centre hotel a real treat</i>	Hizo de este hotel [...] un verdadero placer	Traducción literal no natural en español.
<i>A long weekend</i>	Un largo fin de semana	Esta expresión en inglés se refiere a un «puente», no a un fin de semana largo
<i>Good value for money</i>	Buen valor para el dinero	La TA traduce de manera literal la expresión inglesa. En español existe la expresión equivalente «buena relación calidad-precio».
<i>Makes all the difference</i>	Hace la diferencia	Traducción literal no natural en español.
<i>Was a nice touch</i>	Fue un buen toque	Traducción literal no natural en español.
<i>Very pleased on all fronts</i>	Muy contento en todos los frentes	Traducción literal no natural en español.
<i>I highly recommend</i>	Lo recomiendo altamente	Traducción literal no natural en español.
<i>Finder's fee ¿????</i>	Tarifa de buscador	Traducción literal. En español el concepto más similar sería una «tarifa o comisión de intermediario»
<i>2 thumbs up</i>	2 aprobado	Traducción literal que ocasiona la pérdida total del sentido.
<i>Deffo</i>	Deffo	la TA no lo reconoce y lo deja sin traducir
<i>Freshly squeezed orange juicing machine had left the building</i>	La máquina de jugo de naranja recién exprimida había abandonado el edificio.	Traducción literal que ocasiona la pérdida del sentido.

Se realiza también un uso abusivo de oraciones pasivas y de pasivas reflejas, que en español no son tan frecuentes como en inglés y se ha detectado también un uso y una formación incorrectos de las oraciones de relativo. Además, también se utilizan de manera excesiva las oraciones coordinadas copulativas, que, aunque en inglés son frecuentes, en español, no se usan tan a menudo. Se han expuesto algunos ejemplos en la Tabla 36 siguiente:

Tabla 36. Traducción incorrecta de oraciones pasivas y de las oraciones de relativo. Corpus de hoteles.

Original	TA	Comentario
<b>Uso de pasivas y de pasivas reflejas</b>		
<i>To be waken up early</i>	Ser despertado tan temprano	En español no es tan común el uso de este tipo de estructuras como en inglés, por eso el uso reiterado de estas en el corpus traducido con TA dota de poca naturalidad a estas reseñas.
<i>To be atended</i>	Ser atendido	
<i>Wasn't brought</i>	No fue traído	
<i>Was also used by us</i>	Fue utilizada por nosotros	En estos casos algunos ejemplos de traducciones correctas correspondientes a los ejemplos anteriores serían:
<i>There were robes and slippers provided</i>	Se proporcionaron batas y pantuflas	
<i>They were dealt with</i>	Fueron atendidos	- Que nos despertaran tan temprano
<i>My diet was abandoned after 2 days!</i>	¡Mi dieta fue abandonada después de 2 días!	- Nos atendieron - Nos trajeron - La utilizamos - Nos proporcionaron batas y pantuflas - Se hicieron cargo de - Abandoné mi dieta
<b>Oraciones de relativo</b>		
<i>Every time I have received an excellent reception.</i>	Cada vez <u>que</u> he recibido una excelente recepción	En el original lo que en realidad expresa es «Cada vez he recibido una excelente bienvenida», en este caso al introducir el «que» se cambia el sentido de la oración
<i>Epecially smiley blond lady receptionist which I sadly didn't take name of!</i>	Especialmente sonriente recepcionista rubia <u>que</u> lamentablemente no tomé nombre!	TA forma una estructura incorrecta. En este caso, sería más correcto el uso de « <u>de quien</u> » o « <u>cuyo</u> nombre no apunté»
<i>And it's a room that I cannot wait to leave the second I step into it</i>	Es una habitación <u>que</u> no puedo esperar para salir	TA forma una estructura incorrecta. En este caso, sería más correcto reformular la oración de la manera siguiente «una habitación de donde no puedo esperar a salir» o «una habitación que no puedo esperar a dejar»
<i>The first time we booked a suite, which I would highly recommend,...</i>	La primera vez que reservamos una suite, <u>que</u> recomiendo encarecidamente,...	TA forma una estructura incorrecta. La traducción sería «La primera vez que reservamos una suite, algo que recomiendo encarecidamente,...»

<i>I was very proud on behalf of our company that they had chosen this venue.</i>	me sentí muy orgulloso en nombre de nuestra empresa <u>de que</u> habían elegido este lugar.	El problema viene por la elección del tiempo verbal. En este caso debería ser «de que <u>hubieran</u> elegido». Se ha incluido en esta categoría, porque cuando se dan oraciones de relativo la TA suele presentar problemas a la hora de elegir el tiempo verbal correcto.
<i>That also after I requested it, wasn't brought.</i>	Que también después <u>de que</u> lo solicité, no fue traído	Al igual que en el ejemplo anterior TA forma una estructura incorrecta y utiliza un tiempo verbal incorrecto. En este caso, sería más correcto reformular la oración de la manera siguiente «Que también después <u>haberlo</u> solicitado...»

---

El principal problema de la TA en cuanto a la formación de oraciones de relativo se plantea en el tipo de estructuras del tipo *adv + sujeto + verbo*, por ejemplo, en oraciones como «Every time I have» o «after I requested». En estos casos, la TA suele introducir un «que» innecesario que cambia el sentido de la oración o que directamente no es correcto en esos contextos.

Por otro lado, la TA suele hacer un uso excesivo de oraciones coordinadas copulativas: *Good quality food and both breakfast and dinner* → «y desayuno y cena» o *pool bar and the staff were really friendly and food* → «piscina y el personal era muy amable y la comida era bastante buena». Otro error detectado es la omisión de artículos, que, aunque en inglés sea un recurso frecuentemente utilizado, en español no es correcta en ciertos contextos como los siguientes: *Hotel was spotless* → «(el) Hotel estaba impecable»; *are perfect, architecture/decoration are not to my taste* → «son perfectas, (la) arquitectura / (la) decoración no es de mi gusto»; *Hotel Las Arenas has meet all of our requirements* → (El) Hotel Las Arenas ha cumplido todos nuestros requisitos; *o we stayed in Las Arenas* → «nos quedamos en (el Hotel) Las Arenas».

Por último, se ha observado que, de manera reiterada en este corpus, cuando hay varios adjetivos seguidos en el original en inglés que se traducen al español con TA, esta tiene problemas con el orden de dichos adjetivos y los suele situar en la oración incorrectamente o de manera poco natural: *Comfortable big beds* → «cómodas camas grandes», *nice clean room* → «bonita habitación limpia», *very good attentive staff* → «muy buen personal atento»,

*Nice big pool* → «bonita piscina grande» o *thoughtful little details* → «pequeños detalles pensativos».

### 3.3.2.2. Problemas culturales de localización que afectan a los niveles de calidad de la TA. Corpus hoteles

De manera similar, también han surgido errores de tipo cultural en el corpus de reseñas de hoteles (véase Tabla 37 a continuación):

Tabla 37. Clasificación de la calidad de la TA en función del número y porcentaje de errores culturales. Corpus hoteles.

CALIDAD TA	Nº errores CULTURAL
<b>Nativo</b>	0
<b>Correcta</b>	10
<b>Aceptable</b>	54
<b>Inaceptable</b>	35

En cuanto a los resultados del análisis del corpus de reseñas de hoteles, destaca a primera vista el hecho de que no haya ni un solo caso de reseña nativa que contenga errores culturales. En el otro extremo está el caso de las reseñas aceptables, las más numerosas, como se observa en la Ilustración 28, con un porcentaje del 21 % del total. Es también destacable el hecho de que, en este caso, el porcentaje de reseñas con un nivel inaceptable sea del 14 %, mucho mayor que en el caso de los restaurantes, donde este porcentaje era solo del 2,4 %. Por último, el porcentaje de reseñas con un nivel de calidad correcto es justo igual que en el caso de los restaurantes, un 4 %.

A continuación, la Ilustración 28 refleja estos datos sobre errores de localización culturales que afectan a la calidad de la TA identificados en el corpus de reseñas de hoteles:

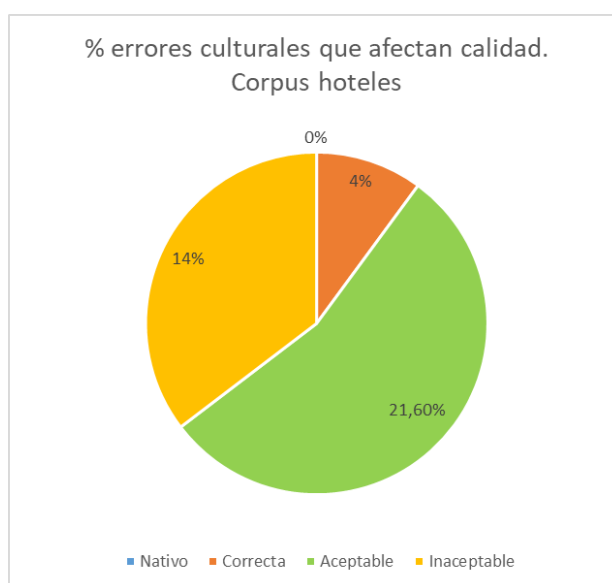


Ilustración 28. Porcentaje de errores culturales que afectan a la calidad de la TA. Corpus hoteles.



En este corpus de reseñas de hoteles no se han dado tantos casos relacionados con comida o platos típicos de la región, ya que los usuarios se centran más en evaluar la calidad de las habitaciones o la estancia en general, por lo que los errores culturales han sido de otro tipo. Por un lado, han surgido errores con la traducción de los nombres de los lugares históricos o turísticos de Valencia. Además, la TA también presenta problemas a la hora de traducir los nombres de las calles y de los barrios u otras denominaciones. Algunos ejemplos de este tipo de problemas se presentan en la Tabla 38, a continuación:

Tabla 38. Errores de localización culturales al traducir nombres de lugares turísticos de la ciudad.

Original	TA	Comentario
<i>Main train station</i>	Estación principal de trenes	El nombre oficial de este lugar y que utilizan los
<i>The Nord train</i>	Del tren Nord	hablantes nativos es «la Estación del Norte»
<i>Main market</i>	Mercado principal	El Mercado central
<i>The green belt park</i>	Parque del cinturón verde	
<i>Park along where river was diverted</i>	El parque por donde se desvió el río	Parque del cauce del río Turia
<i>Maritin serreria</i>	Maritin serreria	Marítim serreria
<i>The Science Centre</i>	Al Centro de Ciencias	
<i>The art city</i>	Ciudad del arte	
<i>Arts and sciences complex</i>	Complejo de artes y ciencias	La Ciudad de las Artes y las Ciencias
<i>The Science/Oceanografic Area</i>	Las áreas de Ciencia	
<i>El corte iglese</i>	El corte iglese	El corte inglés
<i>Russaca</i>	Russaca	Ruzafa/russafa

Es más, la TA llega a traducir una misma denominación de maneras diferentes, a veces incluso en una misma reseña:

- Estación del Norte → estación principal de trenes / tren Nord
- Parque del cauce del río Turia → parque del cinturón verde / el parque por donde se desvió el río
- La Ciudad de las Artes y las Ciencias → Centro de Ciencias / ciudad del arte / complejo de artes y ciencias / las áreas de Ciencia
- El barrio del Cabanyal → área de Cabanyal / el barrio bohemio.

En algunas ocasiones, también han surgido errores al traducir los nombres de platos típicos u otros conceptos relacionados con comida, al igual que pasaba en las reseñas de restaurantes. Se han incluido algunos ejemplos en la Tabla 39 siguiente:

Tabla 39. Errores de localización cultural por traducción incorrecta de ingredientes o platos típicos. Corpus hoteles.

Original	TA	Comentario
<i>Hotdog with chips</i>	<i>Hot dog</i> con patatas fritas	Perrito caliente con patatas fritas
<i>Spanish ham</i>	Jamón español	Jamón / jamón serrano
<i>Cold meats</i>	Carnes frías	fiambres
<i>Restaurant options for "paellas"</i>	Restaurantes para "paellas"	concepto que en ES no se utiliza, sería como mucho «restaurantes especializados en paellas»
<i>Seafood eateries</i>	Restaurantes de mariscos	marisquerías

Por otro lado, en ocasiones la TA no traduce conceptos o expresiones más típicos del inglés, que puede que no se entiendan en español. De la misma manera, los nombres de festividades típicas de Valencia se traducen incorrectamente. Ambos tipos de ejemplos se muestran en la Tabla 40 a continuación:

Tabla 40. Errores de localización culturales por la no traducción de expresiones en inglés o la traducción incorrecta de festividades locales. Corpus de hoteles.

Original	TA	Comentario
<i>4 star hotel b/b</i>	Hotel de 4 estrellas b / b	B/b es un concepto propio de la lengua inglesa que se puede referir al típico <i>bed &amp; breakfast</i>
<i>Without fake American smiling</i>	Sin falsa sonrisa estadounidense	En España por lo general este prejuicio o percepción de los estadounidenses no está tan generalizado
<i>Hop on hop off bus</i>	Autobús Hop on hop off	TA incorrecta, en español existe el término «autobús turístico»

<i>Buck's Fizz and lite English Breakfast</i>	Buck's Fizz y lite English Breakfast	Conceptos propios de la cultura de habla inglesa, pero no de la española. No obstante, en español sí se puede traducir el <i>english breakfast</i> como «desayuno inglés» o, si se refiere al té denominándolo «té <i>English Breakfast</i> »
<i>Boutique hotel</i>	Hotel boutique	Concepto no extendido en la cultura española cuyos hablantes no entenderán
<i>The el Maitre D</i>	El Maitre D	En español se utiliza solo «maitre»
<i>D</i>	Re	El autor de la reseña firma como «D» y la TA lo ha traducido como la nota musical «Re» que en inglés se escribe «D»
<i>A+</i>	A+	En España no se usa este tipo de nomenclatura para las notas
<i>9th Octobe Valencia parade</i>	El noveno desfile de Octubre Valencia	el desfile del 9 de octubre
<i>Las Fallas Festival</i>	El Festival de Las Fallas	Las Fallas

También se han detectado errores en las abreviaturas, los signos o las divisas. A continuación, se muestra la TA y una muestra de cómo deberían traducirse estos conceptos. Se muestran algunos ejemplos en la Tabla 41, a continuación:

Tabla 41. Errores de localización culturales por traducción incorrecta de abreviaturas, signos o divisas. Corpus de hoteles.

Original	TA	Traducción correcta
19hrs	19hrs	19 h
EURO 2-3	EURO 2-3	2-3 EUROS / 2-3 €
140 pounds	140 libras	Se debería haber realizado la conversión a Euros, ya que se habla de precios de un hotel en la ciudad de Valencia donde la divisa no son las Libras sino los Euros
10pm / 9.30 p.m.	10pm / 9.30 p.m.	Los formatos de las horas en español son diferentes al inglés. Por ejemplo: 22:00 y 21:30 respectivamente.
9 inches / 5'3 " / 2 inches / 10 feet /miles / 200 yards	9 pulgadas / 5'3 " / 2 pulgadas / 10 pies / millas /200 yardas	En español se utiliza el sistema métrico, por lo que se deberían haber convertido estas magnitudes.

Por último, en relación con el uso de expresiones no empleadas en la variante del español de España, también aparecen palabras o construcciones más propias de variantes latinoamericanas del español (véase Tabla 42):

Tabla 42. Errores culturales de localización por uso de variantes no propias del español de España. Corpus de hoteles.

Original	TA	Traducción correcta
Smell divine	Huelen divino	Huelen genial
Unpack	Desempacar	Deshacer la maleta
Our block	A una cuadra	En nuestra manzana
How illogical they are	Qué tan ilógicas son	Cómo de ilógicas son
Parking	Estacionamiento	Aparcamiento / parking
It feels good	Se siente bien	Es agradable
The car	El auto	El coche
Husband	Esposo	Marido
Stay	Estadía	Estancia

<i>Tickets</i>	Boletos	Billetes/ tickets
<i>Beautiful</i>	Lindo	Bonito
<i>Laminate flooring</i>	Piso laminado	Suelo laminado
<i>Laptop</i>	Computadora portátil	Ordenador portátil
<i>We had a great time</i>	La pasamos de maravilla	Lo pasamos de maravilla
<i>Orange juice</i>	Jugo de naranja	Zumo de naranja
<i>Additional cost</i>	El costo adicional	El coste adicional
<i>Fridges</i>	Refrigeradores	Neveras
<i>In the morning</i>	En la mañana	Por la mañana

### 3.3.2.3. Causas frecuentes que afectan al nivel de calidad en el corpus de reseñas de hoteles

Por otra parte, tras analizar el número de casos de este tipo de errores que afectan a la calidad de la TA en el corpus de hoteles, se han hallado los datos expuestos en la Tabla 43 siguiente:

Tabla 43. Número problemas que afectan al nivel de calidad de la TA. Corpus hoteles.

<b>CAUSAS QUE AFECTAN A LA CALIDAD DE LA TA</b>	<b>Nº errores</b>
<b>Problemas explicación / duplicidad</b>	14
<b>Problemas calidad motor TA</b>	194
<b>Problemas error en el original</b>	32

La Tabla 43 muestra los tipos de problemas que afectan a la calidad de la TA de las reseñas recopiladas en el corpus de hoteles. A diferencia del corpus de restaurantes, en el de hoteles, los errores de explicación y duplicidad han sido los menos numerosos con un total de solo 14 casos. En esta ocasión, los errores más frecuentes han sido aquellos causados por la calidad del motor de TA, con un total de casos de 194. Por último, en el corpus de hoteles se han producido un total de 32 casos de problemas a causa de errores en las reseñas originales. En la Ilustración 29, a continuación, se ofrecen los datos relativos a los porcentajes que ocupan estos errores:



Ilustración 29. Porcentaje factores que afectan al nivel de calidad de la TA. Corpus hoteles.

Gracias a la Ilustración 29 anterior, se aprecia a simple vista la diferencia entre los tipos de errores detectados en cada corpus. En este caso, la principal diferencia se aprecia en el porcentaje de ítems con problemas de calidad a causa de errores en el original, que en este caso es mucho menor que en el de las reseñas de restaurantes (tan solo 5,6 % en el caso de los hoteles, frente al 86,8 % en el caso de los restaurantes), algo que se podría atribuir al hecho de que, en el caso de los hoteles, no se utiliza tanto vocabulario relativo a conceptos o platos regionales como se llega a utilizar cuando se elabora una reseña sobre un restaurante típico de la zona.

A continuación, se ofrecen los resultados obtenidos en el análisis de los problemas de explicación o duplicidad en el caso del corpus de reseñas de hoteles. Como se ha mencionado anteriormente, en el caso del corpus de reseñas de hoteles, no se han producido muchos casos de problemas de explicación o duplicidad, y estos se han dado principalmente cuando en inglés se utilizaban adjetivos sinónimos que, al ser traducidos con TA a español se han repetido. No obstante, también han surgido casos con las denominaciones del tipo de habitaciones, ya que en inglés existen conceptos como *double room* o *twin room* que en español no se utilizan. En la Tabla 44 se presentan algunos ejemplos:

Tabla 44. Ejemplos de errores de duplicidad o explicación encontrados. Corpus de hoteles.

Original	TA
<u>Excellent</u> location [...] <u>great</u> restaurants	<u>Excelente</u> ubicación [...] <u>excelentes</u> restaurantes
<u>helpful and accommodating</u>	muy <u>servicial</u> y <u>servicial</u>

<u>bartender/waiter</u>	era el <u>camarero</u> / <u>camarero</u>
<u>mix between old and new is a wonderful blend</u>	La <u>mezcla</u> entre lo viejo y lo nuevo es una <u>mezcla</u>
<u>tiny little piece</u>	una <u>pequeña</u> y <u>pequeña</u> pieza
were <u>happy</u> , [...] and a <u>joy</u> to be served by	los demás estaban <u>felices</u> , [...] y <u>felices</u> de ser atendidos
<u>dresser/chest</u>	<u>cómoda/cómoda</u>
<u>double room-twin room</u>	<u>habitación doble-habitación doble</u>
<u>double room</u> to a <u>twin</u>	<u>habitación doble</u> a una <u>doble</u>

Siguiendo por los errores causados por la calidad del motor de TA, que en el caso del corpus de hoteles han sido los más abundantes, se han vuelto a encontrar errores en el uso de verbos como «ser» y «estar» cuando en inglés se utiliza el verbo *to be*, así como errores en el uso del verbo «tener» cuando en inglés se utiliza *to have*, no solo utilizados en relación con la comida, sino también en otros contextos. Se han encontrado los ejemplos detallados en la Tabla 45:

Tabla 45. Traducción incorrecta de los verbos *to be* y *to have*. Corpus hoteles.

Original	TA	Comentario
<b>Uso incorrecto del verbo «tener»</b>		
We <u>had</u> a breakfast	<u>Teníamos</u> un desayuno	<u>Desayunamos</u> / <u>Tomamos</u> un desayuno
I <u>had</u> some champagne	<u>Tuve</u> un poco de champán	<u>Tomé</u> / <u>bebí</u> un poco de champagne
We <u>have</u> light breakfasts at home	<u>Tenemos</u> desayunos ligeros en casa	<u>Desayunamos</u> ligero / <u>Tomamos</u> desayunos ligeros
We <u>had</u> value for money	<u>Teníamos</u> una buena relación calidad-precio	«El hotel <u>ofrecía</u> una buena relación calidad-precio».
I <u>had</u> some complimentary champagne	<u>Tuve</u> un poco de champán	<u>Tomé</u> un poco de champán / <u>Bebí</u> un poco de champán
We <u>have</u> light breakfasts at home	<u>Tenemos</u> desayunos ligeros en casa	<u>Desayunamos</u> ligero / <u>Tomamos</u> desayunos ligeros
We <u>had</u> a hire car	<u>Teníamos</u> un coche de alquiler	La expresión no es incorrecta en sí, no obstante, suena mucho más natural « <u>Alquilamos</u> un coche»

<i>They <u>have</u> weddings and business conferences</i>	<u>Tienen</u> bodas y conferencias de negocios	<u>Celebran</u> bodas y conferencias de negocios
<i>We <u>had</u> free parking</i>	<u>Tuvimos</u> aparcamiento gratuito	<u>Había</u> servicio de aparcamiento gratuito
<i>We <u>had</u> a break</i>	<u>Tuvimos</u> un descanso	<u>Descansamos</u>

---

**Uso incorrecto de los verbos «ser / estar»**

---

<i>The breakfast was standard, <u>wasn't</u> diverse enough in my opinion</i>	El desayuno era estándar, no <u>estaba</u> lo suficientemente diverso en mi opinión.	El desayuno no <u>ERA</u> lo suficientemente variado
<i>It <u>was</u> delicious</i>	<u>Era</u> delicioso	<u>Estaba</u> delicioso
<i>The buffet breakfast <u>was</u> good value</i>	El desayuno buffet <u>era</u> una buena relación calidad-precio	Ha confundido el uso del verbo «ser» con el «tener». El desayuno buffet <u>tenía</u> una buena relación calidad-precio
<i>Breakfast <u>was</u> delicious</i>	El desayuno <u>era</u> delicioso	El desayuno <u>estaba</u> delicioso
<i>Paella <u>was</u> amazing</i>	La paella <u>era</u> increíble	La paella <u>estaba</u> increíble
<i>Not much of a view</i>	No es una gran vista	No <u>tiene/cuenta con</u> una gran vista
<i>Standard rooms which <u>were</u> a decent size</i>	Habitaciones estándar que <u>eran</u> de un tamaño decente	Ha confundido el uso del verbo «ser» con el «tener». «Habitaciones estándar que <u>tenían</u> un tamaño decente»
<i>Balcony which <u>was</u> of good size</i>	Balcón que <u>era</u> de buen tamaño	Ha confundido el uso del verbo «ser» con el «tener». «Balcón que <u>tenía</u> un buen tamaño»
<i>...you will be disappointed.</i>	...se sentirá decepcionado.	...se sentirá decepcionado.
<i>I definitely <u>am</u>.</i>	Definitivamente lo <u>soy</u> .	Definitivamente lo <u>estoy</u> .
<i>The room <u>was</u> pristine</i>	La habitación <u>era</u> impecable	La habitación <u>estaba</u> impecable

---

Otros ejemplos de errores causados por la calidad del sistema de TA son los enumerados en la Tabla 46, a continuación:



Tabla 46. Ejemplos de errores causados por la calidad del motor de TA. Corpus hoteles.

Original	TA	Comentario
<i>Extensive experiences not incurred</i>	No se incurre en experiencias extensas	Traducción literal. Se refiere a que no tuvieron la oportunidad de tener muchas experiencias por la corta duración de la estancia
<i>Easy to go to places</i>	Fácil de caminar a lugares	Fácil de ir a los sitios
<i>To walk anywhere</i>	Caminar en cualquier lugar	Se puede ir caminando a cualquier parte
<i>Massive bed</i>	Cama masiva	Cama enorme
<i>Were woken [...] every day to banging and drilling</i>	Nos despertaban[...] Para golpear y perforar	Traducción literal e incorrecta, se refiere a que despertaron a causa de los sonidos de las obras de la calle
<i>Fresh bathroom</i>	Baño fresco	Baño acabado de limpiar
<i>Made this city centre hotel a real treat</i>	Hizo de este hotel [...] un verdadero placer	hizo que hospedarse en este hotel fuera una maravilla
<i>Cold meats cooked food</i>	Carnes frías, comida cocinada	Fiambres, comida caliente
<i>Bucks fizz with fresh orange</i>	Dinero con naranja fresca	Se refiere al tradicional cóctel valenciano espumoso conocido como «Agua de Valencia» que se elabora con zumo de naranja fresca y champán (entre otros ingredientes) de ahí el <i>bucks fizz</i> que la TA traduce como «dinero».

*Spotlessly clean*

Impecablemente limpias

En español es redundante.  
Bastaría con usar el adjetivo  
«impecables»

Además, se han encontrado errores recurrentes como, por ejemplo, la traducción de *welcoming* cuando se habla del personal del hotel y que, en todos los casos, se ha traducido como «acogedor», lo que es incorrecto: *welcoming staff* → «personal acogedor» o *staff was very welcoming* → «personal fue muy acogedor». Asimismo, la TA también comete errores en la traducción del adjetivo *nice* cuando lo traduce como «encantador» en casos como los siguientes: *nice tapas* → «unas tapas encantadoras» o *nice toiletries* → «artículos de aseo encantadores». Al igual que en el caso de los restaurantes, también se cometen errores en el uso excesivo de oraciones pasivas como en el caso de *to be attended* → «ser atendido», *wasn't brought* → «no fue traído» o *be waken up early* → «ser despertado tan temprano», entre otros ejemplos, así como la traducción literal de expresiones idiomáticas del inglés como *it was a nice touch* → «fue un buen toque» o *contributed to my enjoyment* → «contribuyó al disfrute», entre otras.

En el corpus de reseñas de hoteles también han surgido problemas causados por errores en las reseñas originales y, al igual que en el corpus de restaurantes, estos van desde palabras mal escritas que la TA no es capaz de corregir a oraciones que, por un ligero error, han cambiado totalmente el significado del texto original. Estos errores se detallan en la Tabla 47, a continuación:

Tabla 47. Ejemplos de errores causados por errores en la reseña original en inglés. Corpus hoteles.

Original	TA	Comentario
<i>Quite <u>tierd</u></i>	Bastante tierd	Quite <u>tired</u>
<i>You would <u>kind</u></i>	Que te gustaría tener	You would <u>find</u>
<i>The 9th <u>Octobe</u> Valencia parade</i>	El noveno desfile de Octobe Valencia	The 9th <u>October</u> Valencia parade
<i>Valncia</i>	Valncia	Valencia
<i>The <u>stuff</u> was friendly</i>	Las cosas eran amables	The <u>staff</u> was friendly
<i><u>Grate</u> hotel</i>	Grate hotel	<u>Great</u> hotel

<i>We had a wonderful time in your beautiful <u>count</u></i>	La pasamos de maravilla en tu hermoso conde	We had a wonderful time in your beautiful <u>country</u>
<i>Good <u>matrass</u></i>	Buen matrass	Good <u>matress</u>
<i>Almond <u>desert</u></i>	Desierto de almendras	Almond <u>dessert</u>
<i>Congress for <u>accessi</u> blew cruise destinations</i>	Congreso internacional para el acceso	Congress for <u>accessible</u> cruise destinations → este error parece haberse producido por el corrector, ya que muchas de estas reseñas se escriben desde dispositivos móviles
<i>Which is very <u>ptetty</u> Salmon, Spanish ham, <u>figues</u></i>	Que es muy pesado Salmón, jamón español, <u>estatuas</u>	Which is very <u>pretty</u> Salmon, Spanish ham, <u>figs</u> → del contexto se puede extraer que se refiere a higos y la escritura e inglés de esta palabra es <i>figs</i> . La TA ha corregido incorrectamente la palabra pensando que se refería a <i>figures</i> es decir «figuras» o «estatuas», un significado que es erróneo en este contexto

---

### 3.4. Comparativa de los resultados de los corpus bilingües de reseñas de restaurantes y de hoteles

En un análisis más detallado sobre los tipos de errores que afectan a la calidad de las reseñas, se ha observado que los errores de localización encontrados son de tipo cultural o lingüístico. En este caso, del análisis del corpus de reseñas de restaurantes se han extraído los resultados siguientes (véase Tabla 48):

Tabla 48. Casos de problemas culturales o lingüísticos causados por diferentes tipos de errores. Corpus restaurantes.

Tipo de error	Nº errores	% sobre el	Nº errores	% sobre el
	CULTURALES	total	LINGÜÍSTICOS	total
error en original	11	4,4	14	5,6
Problema calidad	32	12,8	69	27,6
Problema explicación/duplicidad	11	4,4	9	3,6
<b>Nº total</b>	<b>52</b>		<b>92</b>	

Como se aprecia en la Tabla 48, en el corpus de reseñas de restaurantes se han identificado un total de 54 errores de tipo cultural. De estos, 11 se han producido por un error en el original, 32 han surgido por problemas de calidad del sistema de TAN y 11 por los problemas de explicación o duplicidad mencionados en apartados anteriores. Además, se han podido observar un total de 92 errores de tipo lingüístico, es decir 60 casos más de tipo lingüístico que de tipo cultural. De estos casos, 14 han surgido por errores en la reseña original, 69 por problemas de calidad del motor de TA y tan solo 9 se han producido por problemas de explicación o duplicidad.

Por otra parte, los resultados del análisis del corpus de reseñas de hoteles con respecto a los errores de localización culturales y lingüísticos que han causado problemas de diferentes tipos han sido los siguientes (Tabla 49):

Tabla 49. Casos de problemas culturales o lingüísticos causados por diferentes tipos de errores. Corpus de hoteles.

Tipo de error	Nº errores	% sobre el	Nº errores	% sobre el
	CULTURALES	total	LINGÜÍSTICOS	total
error en original	14	5,6	30	12
Problema calidad	93	37,2	163	65,2
Problema explicación/duplicidad	0	0	21	8,4
<b>Nº total</b>	<b>107</b>		<b>214</b>	

La incidencia de errores en el corpus de hoteles ha sido mucho mayor que en el de restaurantes, tanto en el caso de los errores de localización culturales (107 casos) como los de tipo lingüístico (214 casos). Dentro de los problemas de tipo cultural, 14 tienen su origen en errores en el original y 93 en problemas de calidad del motor de TA, sin embargo, no se ha hallado ni un solo caso de errores de explicación o duplicidad que hayan dado lugar a errores de tipo cultural, algo que se explica por la baja carga cultural de las reseñas de hoteles en contraste con la alta incidencia de vocabulario regional en las reseñas de restaurantes. En el caso de los problemas de tipo lingüístico, los errores más frecuentes con diferencia han sido los surgidos por la calidad

del motor de traducción (163 casos), seguidos de lejos por los causados por errores en el original (30 casos) y por los ocasionados por problemas de explicación o duplicidad (21 casos).

Además de los problemas anteriores, han aparecido otros aspectos relevantes tanto en el corpus bilingüe de reseñas de restaurantes como en el de las reseñas de hoteles, que también contribuyen a que la TA no alcance un nivel de calidad nativo y a que surjan problemas en las reseñas traducidas. Un ejemplo son las oraciones de relativo, ya que, en numerosas ocasiones, la TA tiene problemas para formarlas correctamente, lo que da como resultado ejemplos como los detallados en la Tabla 50:

Tabla 50. Errores causados por la traducción incorrecta de oraciones de relativo.

Corpus	Original	TA	Comentario
<b>Oraciones de relativo</b>			
Restaurantes	<i>It was the second time we visited el Palmar</i>	Era la segunda vez visitamos el Palmar	Omisión del «que»
Restaurantes	<i>This is the best restaurant we ate in</i>	es el mejor restaurante que comimos	[...] en que hemos comido
Hoteles	<i>there was no tea in the room, which the other hotel provided</i>	no había té en la habitación, que el otro hotel proporcionó	Formación incorrecta
Hoteles	<i>This is the first time we booked a suite, which I highly recommend</i>	La primera vez que reservamos una suite, que recomiendo encarecidamente	Formación incorrecta

También se han encontrado casos en que los nombres con mayúscula inicial no se traducen porque la TA los toma como nombre propio, aunque no lo sean siempre. Algunos ejemplos se observan en la Tabla 51:

Tabla 51. Errores por detección incorrecta de las mayúsculas.

Corpus	Original	TA	Comentario
<b>Oraciones de relativo</b>			
Restaurantes	<i>Iberian Ham Sampler</i>	Sampler de Jamón Ibérico	Se refiere a una muestra o a una loncha de jamón de prueba

Restaurantes	<i>House Red</i>	House Red	Se refiere al vino tinto de la casa
Restaurantes	<i>a night Apple Hooch</i>	noche Apple Hooch	Se refiere a que tomaron un cóctel Apple Hooch por la noche
Restaurante	<i>Valencia Experience</i>	Valencia Experience	Se refiere a una experiencia en Valencia, no es un nombre propio
Hoteles	<i>Valencia Hospitality</i>	Hospitality Valencia	La TA simplemente invierte el orden de la oración, cuando en realidad se debería haber traducido por algo como «hospitalidad valenciana»
Hoteles	<i>The Old Town</i>	The Old Town	Se refiere al casco antiguo de la ciudad, al haber escrito cada palabra en mayúscula la TA lo toma como un nombre propio
Hoteles	<i>Buffett style</i>	Estilo Buffett	Se refiere al desayuno estilo bufé, al haber escrito la palabra en mayúscula la TA la interpreta como un nombre propio. Lo correcto hubiera sido

			adaptar la escritura al español y ponerlo en minúscula, es decir, «bufé»
Hoteles	<i>Hotel Excellence</i>	Hotel Excellence	La TA lo toma como el nombre del hotel cuando en realidad se refiere a que el hotel fue excelente

Otro aspecto problemático es el uso inconsistente de «tú» o «usted», es decir, en unas reseñas se usa «tú» y en otras «usted», o incluso en una misma reseña se utilizan ambos pronombres. En cuanto a aspectos más visuales, el uso de emoticonos o símbolos también ha planteado problemas a la TAN, que no los ha reconocido y ha optado por no traducirlos o ha traducido el resto de la oración que contenía el emoticono de manera incorrecta. Se han expuesto algunos ejemplos en la Tabla 52 siguiente:

Tabla 52. Errores causados por la emisión de los emoticonos en la TA de las reseñas.

Corpus	Original	TA	Comentario
Restaurantes	<i>And we go out for dinner a lot 😊</i>	Y salimos a cenar mucho	Omisión del emoticono
Restaurantes	<i>Definitely the best paella place in the region 🍷🍴</i>	Definitivamente el mejor lugar de paella en la región.	Omisión del emoticono
Hoteles	<i>We would stay here again 👍</i>	Nos alojaríamos aquí de nuevo	Omisión del emoticono

Por último, es interesante tener en cuenta el uso de los signos de exclamación en las reseñas en español. En el corpus en inglés se ha detectado un elevado número de expresiones con exclamaciones, un recurso que en español no es tan usual. No obstante, al haber sido traducidas con TAN, las reseñas en español también añaden esta elevada cantidad de exclamaciones, lo que resulta en un uso abusivo de este recurso. Véase el siguiente ejemplo extraído del corpus (Tabla 53):

Tabla 53. Errores causados por un uso abusivo de exclamaciones.

Corpus	Original	TA	Comentario
Restaurantes	<i>Exceptional service and fantastic food!!</i>	Servicio excepcional y comida fantástica !!	En español se utilizan también exclamaciones

---

			al principio de la oración y sin espacio entre la palabra y el símbolo
Restaurantes	<i>Both were very good! I ordered half a jug of sangria, unaware that this would do about 4 glasses, but it was nice and we were in no rush to move after a very large lunch!</i>	¡ambas estaban muy buenas! Pedí media jarra de sangría, sin saber que esto serviría para unos 4 vasos, ¡pero fue agradable y no teníamos prisa por movernos después de un almuerzo muy grande!	Uso excesivo de exclamaciones en español
Restaurantes	<i>they are too special and not like anywhere else! [...] they will make you feel home!</i>	¡son demasiado especiales y no como en ningún otro lugar! [...] ¡te harán sentir como en casa!	Uso excesivo de exclamaciones en español
Hoteles	<i>One of the best hotels and you can walk from the hotel to almost everything!! [...] it's a beautiful beach and Valencia is a beautiful city! The people are very nice too!</i>	¡Uno de los mejores hoteles y puedes caminar desde el hotel a casi todo! [...] ¡es una playa hermosa y Valencia es una ciudad hermosa! ¡La gente también es muy amable!	Uso excesivo de exclamaciones en español
Hoteles	<i>you must all know who I am talking about !!</i>	¡¡¡ todos deben saber de quién estoy hablando !!	En este caso, no solo se produce un uso excesivo de exclamaciones en español, sino que además la TA añade tres



			signos de exclamación al inicio cuando en inglés solo aparecen dos
Hoteles	<i>Everything was perfect!</i>	¡Todo era perfecto!	Uso excesivo de exclamaciones en español

---

En el caso de los hoteles, los principales errores que ha cometido la TAN son la traducción incorrecta de los nombres de lugares históricos o turísticos de la ciudad de Valencia. Por otro lado, en el caso de los restaurantes, que en este caso eran restaurantes de comida típica valenciana, se ha detectado un gran número de errores en la traducción de nombres ingredientes o platos regionales. Dado que en ambos corpus algunos de los errores que más han afectado a la calidad de las reseñas han sido de tipo cultural, sería posible concluir que la TAN todavía tiene problemas a la hora de traducir elementos con una alta carga cultural, por lo que sería de la gran importancia poseer los contenidos, teniendo en cuenta los principios de la localización.

### 3.5. Comparativa de la naturalidad y autenticidad en las reseñas escritas originalmente en español

Este análisis se basa en la evaluación de dos corpus monolingües, uno de reseñas de restaurantes y otro de reseñas de hoteles escritas originalmente en español, para contrastar las características comunes de las opiniones escritas en español con las opiniones escritas originalmente en inglés y traducidas al español mediante TAN directamente en la plataforma de TripAdvisor. Algunas de las características comparadas han sido la extensión de las reseñas, o la frecuencia con que aparecen ciertas palabras en comparación con el corpus bilingüe de reseñas traducidas con TAN. El objetivo de este análisis ha sido identificar si las reseñas traducidas al español con TAN presentan las mismas características que las reseñas escritas originalmente en español.

Para llevar a cabo este análisis, se han recopilado dos corpus de reseñas escritas por hablantes nativos de español de España, que posteriormente se han procesado con el *software* para análisis de corpus WordSmith. Para ello, se ha utilizado una *stoplist* de elaboración propia que contiene aquellas palabras irrelevantes para el análisis (conjunciones, preposiciones o artículos, entre otras palabras) para extraer una lista de palabras frecuentes o *wordlist*. Estas *stoplists* son necesarias, ya que, si estas no se hubieran incluido, los análisis hubieran producido datos erróneos al tener en cuenta palabras irrelevantes que podrían afectar los porcentajes utilizados

para medir la ocurrencia de las palabras más frecuentes de cada corpus. En las secciones siguientes, se exponen los resultados de dichos análisis.

### 3.5.1. Comparativa sobre la naturalidad y la autenticidad en el corpus de reseñas sobre restaurantes. Corpus monolingüe y bilingüe

En los siguientes apartados, se presentan los resultados relativos al análisis del corpus monolingüe de reseñas de restaurantes escritas originalmente en español, en comparación con el corpus bilingüe de reseñas traducidas con TA al español. Así se ha obtenido una visión general de en qué puntos difieren y en cuáles se asemejan las reseñas traducidas al español con TA y las reseñas escritas en español por hablantes nativos del *locale* hablante de español de España. De esta manera, mediante el uso de listas de frecuencias extraídas con WordSmith, ha sido posible observar si los hablantes de inglés tienen en cuenta los mismos aspectos que los hablantes de español cuando estos elaboran sus reseñas.

#### 3.5.1.1. Datos sobre palabras más frecuentes en el corpus bilingüe traducido con TA. Restaurantes

La primera parte de este análisis se ha centrado en evaluar las palabras con más ocurrencias en las reseñas traducidas al español con TA neuronal, para luego contrastar estos resultados con la frecuencia en que estas mismas palabras aparecen en las reseñas escritas originalmente en español. El resultado del corpus de restaurantes se expone en la Tabla 54, a continuación:

Tabla 54. Palabras más frecuentes en reseñas de restaurantes traducidas con TA al ES.

Frecuencia ES Nativo		Frecuencia ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
COMIDA	0,58	COMIDA	1,21
RESTAURANTE	0,43	RESTAURANTE	0,77
VALENCIA	0,17	VALENCIA	0,59
PROPIETARIO	0	PROPIETARIO	0,53
TAPAS	0,04	TAPAS	0,44
VINO	0,09	VINO	0,42
VINOS	0,05	VINOS	0,19
CROQUETAS	0,03	CROQUETAS	0,25
ANFITRIÓN	0	ANFITRIÓN	0,24
EXPERIENCIA	0,10	EXPERIENCIA	0
MENÚ	0,11	MENÚ	0,22
DUEÑO	0,02	DUEÑO	0,21
NOCHE	0	NOCHE	0,19
VACACIONES	0	VACACIONES	0,04
RECOMENDACIONES	0	RECOMENDACIONES	0,14
MARISCOS	0,02	MARISCOS	0,13

VIAJE	0	VIAJE	0,13
ESPAÑA	0	ESPAÑA	0,11
INGLÉS	0	INGLÉS	0,11
ESPAÑOL	0	ESPAÑOL	0,09
JOYA	0	JOYA	0,07

La palabra que más llama la atención a primera vista es «comida», palabra mucho más frecuente en el corpus de reseñas traducidas al español con TA, donde ha tenido una ocurrencia del 1,21 %, frente al 0,58 % de esta misma palabra en el corpus de reseñas escritas originalmente en español. También con relación a la comida, la palabra «croquetas» aparece un 0,25 % en el corpus de TA y solo un 0,03 % en el corpus de reseñas escritas originalmente en español. Esto se puede deber al hecho de que los españoles y los valencianos en concreto, si van a una arrocería no suelen pedir tapas, como pueden ser las croquetas, a no ser que entren en algún menú, sino que se centran más en los arroces. Por otro lado, la palabra «mariscos» tiene una ocurrencia del 0,13 % en las reseñas traducidas con TA, no obstante, en las reseñas escritas originalmente en español, esta misma palabra solo tiene una ocurrencia del 0,02 %, algo que se explica por el hecho de que la palabra «mariscos», en plural, no se usa tan frecuentemente en español, puesto que se prefiere la palabra en singular «marisco».

Llama también la atención que la palabra «tapas» cuenta con un 0,44 % de ocurrencias en el corpus traducido con TA, frente al 0,04 % con que esta aparece en el corpus de reseñas originales en español. Asimismo, los vinos se tienen más en cuenta en el corpus bilingüe, ya que las palabras «vino» y «vinos» cuentan con una frecuencia del 0,42 % y 0,19 % respectivamente en dicho corpus bilingüe, pero solo con un 0,09 % y 0,05 % respectivamente en el corpus de reseñas escritas en español. También salta a la vista la palabra «joya», que ha aparecido con una frecuencia del 0,07 % en el corpus de reseñas traducidas con TA para describir los restaurantes («el restaurante era una joya»), una expresión que además de poco común en español, no es natural para los hablantes nativos de este idioma, algo que se demuestra por el hecho de que esta palabra no haya aparecido en el corpus de reseñas escritas en español.

Curiosamente, la palabra «restaurante» solo cuenta con un 0,43 % de ocurrencias en el corpus monolingüe de reseñas en español, frente al 0,77 % con que aparece en el de reseñas traducidas con TA. Esto podría explicarse porque, en español, a estos restaurantes especializados en arroces, se les llama más frecuentemente «arrocéricas». De igual forma la palabra «comida» aparece solo un 0,58 % de ocasiones en las reseñas escritas por hablantes nativos de español y un 1,21 % de veces en las reseñas traducidas mediante TA.

También se ha observado que los hablantes de inglés parecen dar más importancia a valorar aspectos relacionados con las personas que regentan o que trabajan en el restaurante. Por ejemplo, la palabra «propietario» cuenta con un porcentaje de ocurrencias del 0,53 % en el corpus de reseñas traducidas con TA, pero en el corpus de reseñas en español, esta ni siquiera aparece. De manera similar, la palabra «anfitrión» aparece en un 0,24 % de ocasiones en el corpus de reseñas traducidas con TA, pero esta palabra tampoco aparece en el corpus de reseñas redactadas originalmente en español. Por último, la palabra «dueño» cuenta con un porcentaje de ocurrencias del 0,21 % en las reseñas traducidas al español con TA, frente al solo 0,02 % con que esta aparece en el corpus de reseñas escritas en español.

Por otra parte, también se han encontrado unos porcentajes relevantes de ocurrencias de palabras relacionadas con los viajes. De hecho, la palabra «viaje» cuenta con un porcentaje de ocurrencias del 0,13 % en el corpus de reseñas traducidas con TA, pero esta misma palabra no aparece en el corpus de reseñas redactadas originalmente en español. Del mismo modo, el porcentaje de ocurrencias de la palabra «vacaciones» es del 0,04 % en las reseñas traducidas con TA, y del 0 % en las reseñas escritas en español. Asimismo, los hablantes nativos de inglés parecen tomar más en cuenta las recomendaciones cuando viajan que los españoles, ya que la palabra «recomendaciones» aparece un 0,14 % en las reseñas traducidas con TA, pero no aparece ni una sola vez en el corpus de reseñas escritas en español. Un posible motivo para esto es que los españoles, y más en concreto, los valencianos ya conocen la zona, por lo que puede que no necesiten tanto estas recomendaciones.

Se ha centrado la atención también en los adjetivos que aparecen en cada corpus y la frecuencia con que estos aparecen. Los resultados se exponen en la Tabla 55, a continuación:

*Tabla 55. Adjetivos más frecuentes en reseñas de restaurantes traducidas con TA al ES.*

Frecuencia ES Nativo		Frecuencia ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
EXCELENTE	0,33	EXCELENTE	0,49
EXCELENTES	0,02	EXCELENTES	0,11
INCREÍBLE	0,02	INCREÍBLE	0,28
AMABLE	0,17	AMABLE	0,33
AGRADABLE	0,16	AGRADABLE	0,23
DELICIOSO	0,06	DELICIOSO	0,14
DELICIOSA	0,02	DELICIOSA	0,21
SERVICIAL	0	SERVICIAL	0,13
FANTÁSTICO	0,02	FANTÁSTICO	0,12
FRESCO	0	FRESCO	0,08
FRESCOS	0	FRESCOS	0,12
ENCANTADOR	0,02	ENCANTADOR	0,11

ENCANTADORA	0,02	ENCANTADORA	0,03
RECOMENDABLE	0,02	RECOMENDABLE	0,09
AUNTÉNTICA	0	AUTÉNTICA	0,08

El adjetivo con mayor porcentaje de frecuencia en el corpus de reseñas traducidas con TA ha sido «excelente» y «excelentes», con un 0,49 % y 0,11 % respectivamente. Estas mismas palabras han aparecido un 0,33 % y un 0,02 % respectivamente en el corpus de reseñas en español. Igualmente, la palabra «increíble» ha aparecido con una frecuencia del 0,28 % en el corpus de reseñas traducidas con TA y solo un 0,02 % en el caso de las reseñas escritas originalmente en español. También llama la atención la utilización de adjetivos para describir la comida, como «delicioso» o «deliciosa» los cuales aparecen con mucha más frecuencia en el corpus de reseñas traducidas con TA, 0,14 % y 0,21 % respectivamente, que en el de reseñas traducidas al español, 0,06 % y 0,02 % respectivamente.

Llama la atención que la TA traduce el adjetivo *helpful* como «servicial» en todas las ocasiones que este ha aparecido, de esta manera, en el corpus de reseñas traducidas con TA dicha palabra aparece un 0,13 %. Este adjetivo parece no ser tan utilizado en español, ya que no ha aparecido ninguna vez en todo el corpus de reseñas escritas en español. Esto puede deberse al hecho de que, cuando en español se quiere expresar algo equivalente a *helpful* en el sentido de que alguien fue de ayuda, esto no se hace utilizando adjetivos sino otro tipo de expresiones como, por ejemplo, «me ayudó mucho».

Por último, otro dato que indica que los hablantes de inglés no tienen en cuenta los mismos aspectos que los hablantes de español de España a la hora de redactar sus reseñas es el hecho de que la frescura de los alimentos se tenga mucho más en cuenta en las reseñas traducidas al español, algo que se denota por el porcentaje en que aparecen las palabras «fresco» y «frescos» en el corpus bilingüe, 0,08 % y 0,12 % respectivamente, frente al porcentaje de ocurrencias en el corpus monolingüe en español, 0 % en ambos casos. Es también el caso de «auténtica» refiriéndose a la paella, que aparece con una frecuencia del 0,08 % en las reseñas traducidas con TA, pero que no se menciona en el corpus en español.

### 3.5.1.2. Datos sobre palabras más frecuentes en el corpus monolingüe. Restaurantes

Tras haber analizado el corpus bilingüe de reseñas traducidas con TA, posteriormente, se ha elaborado también una lista de frecuencias de las palabras más comunes en el corpus de reseñas redactadas originalmente en español. Así, se ha comparado el porcentaje de frecuencias de las palabras de este corpus con el porcentaje de ocurrencias de estas mismas palabras en el corpus

bilingüe de reseñas escritas en inglés y traducidas al español con TA neuronal. Los resultados se presentan en la Tabla 56, a continuación:

Tabla 56. Palabras más frecuentes en reseñas de restaurantes corpus monolingüe.

Frecuencia ES Nativo		Frecuencia ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
ARROZ	1,26	ARROZ	0,16
ARROCES	0,32	ARROCES	0,02
PAELLA	0,66	PAELLA	0,40
SERVICIO	0,62	SERVICIO	0,45
CALIDAD	0,53	CALIDAD	0,20
PRECIO	0,51	PRECIO	0,13
CARABINEROS	0,40	CARABINEROS	0,02
TRATO	0,40	TRATO	0
ENTRANTES	0,38	ENTRANTES	0,02
ATENCIÓN	0,33	ATENCIÓN	0,06
MELOSO	0,29	MELOSO	0
PALMAR	0,32	PALMAR	0,07
ZONA	0,23	ZONA	0,07
SABOR	0,22	SABOR	0,09
FAMILIA	0,17	FAMILIA	0,08
CLOCHINAS	0,16	CLOCHINAS	0
BOGAVANTE	0,14	BOGAVANTE	0
MARISCO	0,14	MARISCO	0,07
ALLIPEBRE	0,08	ALLIPEBRE	0
TERRAZA	0,12	TERRAZA	0

En el caso del corpus monolingüe en español, se observa que las palabras «arroz» / «arroces» han tenido la mayor incidencia con un 1,26 % y un 0,32 % respectivamente. En el corpus de reseñas traducidas al español con TA, el caso es justo el opuesto, ya que solo se han producido un 0,16 % de la palabra «arroz» y un 0,02 % de su plural, «arroces». De nuevo, algo que muestra que los integrantes del *locale* español consideran otros aspectos a la hora de redactar y valorar las reseñas que los angloparlantes. De igual forma, otras palabras relacionadas con la gastronomía y sobre todo con la paella tienen mayor incidencia en el corpus monolingüe de reseñas redactadas originalmente en español. Es el caso de «paella» (0,66 % en el corpus de reseñas redactadas en español y 0,40 % en el de reseñas traducidas con TA) o de «meloso» (0,66 % en el corpus de reseñas en español y ninguna ocurrencia en el de reseñas traducidas con TA). De hecho, en alguna ocasión se tradujo incorrectamente el plato de «arroz meloso» como «paella espesa», algo que jamás diría una persona local.

De manera similar, la palabra «carabineros» tiene un mayor porcentaje de ocurrencias en el corpus de reseñas en español (0,40 %) que en el de reseñas traducidas con TA (0,02 %). La

palabra «bogavante» es otro claro ejemplo de la variedad de denominaciones que un nativo de español utiliza normalmente para referirse al marisco, así, se observa que en el corpus de reseñas escritas en español la incidencia de esta palabra es del 0,14 %, cuando en el corpus traducido con TA no se han detectado casos de esta palabra. Esta similitud en el número de ocurrencias se puede producir porque, a diferencia del caso de los carabineros o de los bogavantes (palabras mucho más frecuentes en el corpus monolingüe), la palabra «marisco» es mucho más general y es ampliamente conocida por ambos *locales*. Lo mismo pasa con la palabra «clóchinas», que en el corpus de reseñas escritas en español ha aparecido un 0,16 %, pero en el de TA no ha tenido casos. Es más, esta palabra es todavía más local de la Comunidad Valenciana, ya que en otras partes de España a este tipo de molusco se le denomina «mejillón», aunque no sean exactamente lo mismo.

Examinando las palabras relacionadas con la gastronomía valenciana, también se ha encontrado un porcentaje del 0,08 % de apariciones de la palabra «allipebre» en el corpus de reseñas escritas en español, pero ni un solo caso de esta palabra en el corpus traducido con TA. Esto se atribuye al hecho de que el allipebre no es uno de los platos más conocidos en el extranjero, y a que, por lo general, cuando personas de otros países vienen a Valencia, se decantan por probar la paella, las tapas u otros platos más conocidos. Por lo tanto, se podría concluir que estos conceptos están muy marcados culturalmente y que, como no existen en el *locale* de los autores angloparlantes, estos no los utilizan. Además, la palabra «marisco», en singular, se ha utilizado con un porcentaje del 0,14 %, pero en el corpus de reseñas traducidas con TA, esta se ha utilizado en un 0,07 % de ocasiones. Esto prueba la afirmación que se hacía en el análisis anterior, de que la utilización del plural de esta palabra, que se utiliza mucho en el corpus de reseñas traducidas con TA, pero que no presentó ningún caso en el corpus de reseñas en español, no es nada frecuente entre los hablantes nativos de español.

Los hispanohablantes también parecen valorar más el «trato» (0,40 % en el corpus español, pero ningún caso en el de TA) y la «atención» (0,33 % en el corpus español, solo 0,06 % en el de TA) recibidos, ya que ambas palabras cuentan con un porcentaje mucho más alto de ocurrencias en el corpus monolingüe. Asimismo, las palabras relacionadas con la compañía con quien se visita el restaurante también han sido frecuentes en el corpus de reseñas escritas en español, por ejemplo, la palabra «familia» aparece un 0,17 % en el corpus de reseñas en español y 0,08 % en el de reseñas traducidas con TA.

El concepto de «calidad» se ha valorado en un porcentaje del 0,53 % en el corpus español, frente al 0,20 % con que este aparece en el corpus bilingüe, algo que puede deberse al hecho de que,

por lo general, al tratarse de alimentos que no se encuentran en los *locales* angloparlantes, los turistas no suelen disponer del criterio para evaluar si se trata de ingredientes de calidad. Por último, los españoles valoran más el que el local disponga de terraza, ya que la palabra «terrace» cuenta con una ocurrencia del 0,12 % en el corpus de reseñas en español, pero no se da un porcentaje representativo de esta palabra en el corpus de reseñas traducidas con TA.

Al igual que en el corpus de reseñas traducidas con TA, en este análisis también se ha prestado atención a los adjetivos utilizados en el corpus monolingüe de reseñas escritas en español. Los resultados se detallan en la Tabla 57, a continuación:

*Tabla 57. Adjetivos más frecuentes en reseñas de restaurantes corpus monolingüe.*

Frecuencia ES Nativo		Frecuencia ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
BUENA	0,48	BUENA	0,25
BUENO	0,29	BUENO	0,14
ACOGEDOR	0,17	ACOGEDOR	0,09
INMEJORABLE	0,10	INMEJORABLE	0
PROFESIONAL	0,10	PROFESIONAL	0,03
BUENÍSIMO	0,09	BUENÍSIMO	0
BUENÍSIMA	0,06	BUENÍSIMA	0
EXQUISITO	0,11	EXQUISITO	0
EXQUISITOS	0,05	EXQUISITOS	0
EXQUISITA	0,09	EXQUISITA	0
TRANQUILO	0,09	TRANQUILO	0,01
RICO	0,07	RICO	0,01
LIMPIO	0,06	LIMPIO	0
GENIAL	0,14	GENIAL	0,09
PERFECTO	0,14	PERFECTO	0,08

Una de las observaciones iniciales que destacan a primera vista es que, en las reseñas en español, no se ha utilizado una gran variedad de adjetivos para describir la comida. Así, aunque en inglés se utilizaban adjetivos como «excelente», «deliciosa» o «fantástica», los nativos de español se decantan más por el uso de adjetivos como «bueno / buena», «buenísimo / buenísima» o «rico / rica». De esta manera, «buena» y «bueno» se han utilizado con una frecuencia del 0,48 % y 0,29 % respectivamente en el corpus de reseñas en español, pero solo un 0,25 % y un 0,14 respectivamente en el corpus de reseñas traducidas con TA. En el caso de los superlativos «buenísimo» y «buenísima», en el corpus de reseñas escritas en español estos se han producido en un 0,09 % y 0,06 % respectivamente, sin embargo, en el corpus de reseñas traducidas con TA no ha habido ni un solo caso. «Rico» aparece un 0,07 % de ocasiones en el corpus de reseñas en español, pero solo un 0,01 % en el de reseñas traducidas con TA. Por último, refiriéndose también a la comida, se han encontrado los adjetivos «exquisito»,



«exquisita» y «exquisitos», con un porcentaje de 0,11 %, 0,09 % y 0,05 % respectivamente en el corpus de reseñas en español, pero sin ningún caso en el corpus de reseñas traducidas con TA.

Otra categoría de adjetivos que no aparecía entre las más mencionadas en el corpus bilingüe de reseñas traducidas con TA es la de los adjetivos que describen el restaurante. Así, encontramos adjetivos como «acogedor», con un 0,17 % en el corpus de reseñas en español y con 0,09 % en el corpus de TA; «tranquilo», con un 0,09 % en el corpus de reseñas en español y con 0,01 % en el corpus de TA; o «limpio», con un 0,06 % en el corpus de reseñas en español, pero sin ningún caso en el corpus de TA.

### 3.5.1.3. Datos sobre palabras con frecuencia similar en ambos corpus. Restaurantes

Este análisis también se ha centrado en determinar qué palabras han tenido una frecuencia similar en el corpus bilingüe y en el monolingüe, para evaluar si existen puntos en común entre las traducciones de reseñas realizadas utilizando un motor de TA y las reseñas escritas por hablantes nativos de español. En primer lugar, se van a presentar los resultados del corpus de restaurantes (Tabla 58).

Tabla 58. Palabras con frecuencia similar en ambos corpus. Corpus restaurantes.

Frecuencia ES Nativo		Frecuencia ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
LOCAL	0,19	LOCAL	0,10
AMBIENTE	0,10	AMBIENTE	0,19
POSTRE	0,15	POSTRE	0,13
PLATO	0,18	PLATO	0,18
RESERVA	0,08	RESERVA	0,05
CANTIDAD	0,07	CANTIDAD	0,01
GUSTO	0,07	GUSTO	0,01
PERSONAS	0,07	PERSONAS	0,06
RACIONES	0,07	RACIONES	0,01
RACIÓN	0,07	RACIÓN	0
VERDURAS	0,07	VERDURAS	0,02
CUMPLEAÑOS	0,06	CUMPLEAÑOS	0,03

En primer lugar, la palabra «local» ha tenido un porcentaje similar de ocurrencias en ambos corpus (0,19 % en el monolingüe de reseñas en español y 0,10 % en el bilingüe de reseñas traducidas con TA). Asimismo, ambas culturas parecen dar una importancia similar al ambiente que haya en el restaurante, puesto que la palabra «ambiente» aparece un 0,10 % en el corpus de reseñas en español y un 0,19 en el de reseñas traducidas con TA).

Aunque anteriormente se ha observado que en el corpus de reseñas en español se utilizaban más frecuentemente palabras como «familia», en ambos corpus encontramos palabras como

«personas» (0,07 % en el monolingüe y 0,06 % en el bilingüe). Ambas culturas parecen tener bastante interés en saber las cantidades de comida que se sirven en los restaurantes puesto que las palabras «cantidad» (0,07 % en el monolingüe y 0,01 % en el bilingüe) y «ración/raciones» (0,07 % en el monolingüe y 0,01 % en el bilingüe) aparecen repetidamente con porcentajes muy similares en ambos corpus.

Para finalizar este análisis también se han examinado los adjetivos que han aparecido con frecuencias similares en ambos corpus. Estos se exponen en la Tabla 59, a continuación:

Tabla 59. Adjetivos con frecuencia similar en ambos corpus. Corpus restaurantes.

Frecuencia ES Nativo		Frecuencia ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
ATENTO	0,14	ATENTO	0,10
MEJOR	0,36	MEJOR	0,32
PEQUEÑO	0,10	PEQUEÑO	0,08
EXCEPCIONAL	0,06	EXCEPCIONAL	0,06
RAZONABLE	0,06	RAZONABLE	0,06
CASERO	0,02	CASERO	0,05

Algo que salta a la vista es que ambas culturas valoran que la comida servida sea casera, ya que el adjetivo «casero» cuenta con un 0,02 % de ocurrencias en el corpus monolingüe de reseñas en español y un 0,05 % en el de reseñas traducidas con TA. Asimismo, ambas culturas han valorado el trato atento del personal y han utilizado adjetivos como «pequeño», «excepcional» o «razonable» de manera similar, aunque sería necesario observar si estos se han utilizado en los mismos contextos.

### 3.5.2. Comparativa sobre la naturalidad y la autenticidad en el corpus de reseñas sobre hoteles. Corpus monolingüe y bilingüe

A continuación, se exponen los resultados del análisis del corpus monolingüe de reseñas de hoteles en comparación con el corpus bilingüe. Al igual que en el análisis de las reseñas de restaurantes, la elaboración de listas de frecuencias mediante el *software* WordSmith, ha permitido observar las diferencias y similitudes que presentan los contenidos traducidos con TA y los contenidos elaborados por hablantes nativos de español. Además, también se ha analizado qué aspectos tienen en cuenta los hablantes de inglés y en cuáles se centran los hablantes de español cuando escriben sus reseñas sobre hoteles.

3.5.2.1. *Datos sobre palabras más frecuentes en el corpus bilingüe traducido con TA.  
Hoteles*

En la primera parte de este análisis se han evaluado las palabras que han aparecido con más frecuencia en las reseñas de hoteles traducidas al español con TA neuronal. A continuación, se presentan los resultados de este mismo análisis, esta vez con las reseñas de hoteles elaboradas por hablantes nativos de español de España. La Tabla 60, a continuación, muestra dichos resultados:

Tabla 60. *Palabras más frecuentes en reseñas de hoteles traducidas con TA al ES.*

Frecuencia ES Nativo		Frecuencia en ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
PLAYA	0,19	PLAYA	0,32
PISCINA	0,18	PISCINA	0,31
RESTAURANTE	0,15	RESTAURANTE	0,28
CIUDAD	0,14	CIUDAD	0,38
UBICACIÓN	0,02	UBICACIÓN	0,24
COMIDA	0,06	COMIDA	0,22
ÁREA	0	ÁREA	0,11
ÁREAS	0	ÁREAS	0,04
VIAJE	0,03	VIAJE	0,10
AUTOBÚS	0	AUTOBÚS	0,08
BALCÓN	0,02	BALCÓN	0,08
TAXI	0	TAXI	0,08
CASCO	0	CASCO	0,07
ESTACIONAMIENTO	0	ESTACIONAMIENTO	0,07
DISTANCIA	0,02	DISTANCIA	0,06
MINUTOS	0,05	MINUTOS	0,19
PROBLEMA	0,08	PROBLEMA	0,11
HUÉSPEDES	0	HUÉSPEDES	0,07
AEROPUERTO	0	AEROPUERTO	0,03
BICICLETAS	0	BICICLETAS	0,05

En el corpus de reseñas traducidas al español con TA, en primer lugar, se encuentran las palabras «playa» y «piscina» con unos porcentajes del 0,32 % y el 0,31 % respectivamente, casi el doble que en el corpus de reseñas escritas originalmente en español, donde estas aparecen solamente en un 0,19 % y 0,18 % de ocasiones. Por lo que parece, en las reseñas que se han traducido con TA se ha dado una mayor importancia a valorar la zona donde se encuentra el hotel, ya que se han encontrado múltiples palabras referentes a este tema, como «ubicación» (0,24 % en el corpus de reseñas traducidas con TA y solo 0,02 % en el corpus de reseñas en español), «área» (0,11 % en el corpus de reseñas traducidas con TA, pero sin ocurrencias en el corpus de reseñas en español), «distancia» (0,06 % en el corpus de reseñas traducidas con TA y solo 0,02 % en el

corpus de reseñas en español) o «minutos» (0,19 % en el corpus de reseñas traducidas con TA y solo 0,05 % en el corpus de reseñas en español). En relación con las valoraciones de la ciudad y de sus atractivos turísticos, la palabra «ciudad» se han encontrado un 0,38 % en el corpus de TA y solo un 0,14 % en el corpus de reseñas en español y la palabra «casco», referida al casco antiguo, aparece en un 0,07 % de ocasiones en el corpus de TA, pero no hay casos en el corpus de reseñas en español.

También salta a la vista que los usuarios no españoles dan mucha importancia a que haya un restaurante en el hotel ya que, en el corpus de TA, la palabra «restaurante» aparece en un porcentaje del 0,28 %, mientras que en corpus de reseñas escritas en español el porcentaje fue de tan solo 0,19 %. En relación con la gastronomía, la palabra «comida» aparece en un 0,22 % de ocasiones en el corpus de traducciones con TA frente a las 0,06 % que aparece en el corpus de reseñas escritas en español.

Un dato curioso es que, en el corpus de TA, aparece mucha más variedad de palabras relacionadas con el transporte público que, además, tienen más ocurrencias que en el corpus de reseñas escritas en español. Estas palabras son «autobús» (0,08 % en el corpus de TA y 0 casos en el corpus de reseñas en español), «taxi» (0,08 % en el corpus de TA y 0 casos en el corpus de reseñas en español) o «bicicletas» (0,05 % en el corpus de TA y 0 casos en el corpus de reseñas en español), que seguramente se refiera a bicicletas de alquiler. Por último, es interesante destacar que la palabra «aeropuerto» solo aparece en el corpus de reseñas traducidas con TA, ya que, es mucho más frecuente que las personas extranjeras viajen en avión a otro país a que lo hagan los residentes de un mismo país dentro de su territorio.

Al igual que con el corpus de reseñas de restaurantes, para las reseñas de hoteles también se han examinado los adjetivos más recurrentes tanto en el corpus bilingüe como en el monolingüe de reseñas escritas originalmente en español. En la Tabla 61, a continuación, se detallan los adjetivos más frecuentes en el corpus de reseñas traducidas con TA:

*Tabla 61. Adjetivos más frecuentes en reseñas de hoteles traducidas con TA al ES.*

Frecuencia ES Nativo		Frecuencia en ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
EXCELENTE	0,22	EXCELENTE	0,46
AMABLE	0,26	AMABLE	0,36
AGRADABLE	0,13	AGRADABLE	0,23
POCO	0,20	POCO	0,20
CERCA	0,19	CERCA	0,24
CÓMODA	0,08	CÓMODA	0,12
CÓMODAS	0,06	CÓMODAS	0,07

FÁCIL	0,03	FÁCIL	0,10
HERMOSA	0	HERMOSA	0,09
HERMOSAS	0	HERMOSAS	0,01
ENCANTADOR	0,01	ENCANTADOR	0,08
ENCANTADORA	0	ENCANTADORA	0,06
COMPLEJO	0	COMPLEJO	0,05
NUEVO	0.04	NUEVO	0.09
HERMOSO	0	HERMOSO	0,04

Como se observa en la Tabla 61 al igual que en el corpus de restaurantes, el adjetivo más frecuente en el corpus de reseñas de hoteles traducidas al español mediante TA es «excelente» con un porcentaje de ocurrencias del 0,46 % frente al 0,22 % del corpus de reseñas en español. De manera similar, los adjetivos «amable» y «agradable», utilizados para describir el hotel, cuentan con un 0,36 % y un 0,23 % de ocurrencias respectivamente en el corpus de TA, pero solo tienen un 0,26 % y un 0,13 % respectivamente en el corpus de reseñas en español.

Otros adjetivos utilizados para describir el hotel y sus habitaciones en el corpus de reseñas traducidas con TA han sido, «hermoso/hermosa/hermosas», con un porcentaje de ocurrencias del 0,04 %, 0,09 % y 0,01 % respectivamente y ningún caso en el corpus de reseñas escritas en español. Asimismo, las palabras «encantador/encantadora» cuentan con un porcentaje de frecuencias del 0,08 % y el 0.06 % respectivamente en el corpus bilingüe de reseñas traducidas con TA, pero solo un 0,01 % de casos de «encantador» y ningún caso de «encantadora» en el corpus de reseñas en español. También se ha calificado la comodidad de las camas con adjetivos como «cómoda/cómodas», con un 0,12 % y un 0,07 % de ocurrencias en el corpus de reseñas traducidas con TA respectivamente, pero un 0,08 % y un 0,06 % respectivamente en el corpus de reseñas escritas en español.

Por último, en relación con la afirmación hecha anteriormente sobre que parece que se han valorado más los aspectos relativos a la ubicación de los hoteles en el corpus de reseñas traducidas con TA que en el de reseñas escritas originalmente en español, se ha encontrado un porcentaje relevante del uso del adjetivo «cerca», con un porcentaje del 0,24 % en el corpus de reseñas traducidas con TA y un 0,19 % en el corpus de reseñas escritas originalmente en español.

### 3.5.2.2. *Datos sobre palabras más frecuentes en el corpus monolingüe. Hoteles*

Tras examinar las palabras más recurrentes en el corpus bilingüe de reseñas de hoteles traducidas mediante TA, el siguiente análisis se centró en encontrar las palabras más frecuentes en el corpus monolingüe de reseñas escritas originalmente en español. Los resultados de dichas palabras y sus frecuencias se detallan en la Tabla 62, a continuación:

Tabla 62. Palabras más frecuentes en reseñas de hoteles corpus monolingüe.

Frecuencia ES Nativo		Frecuencia en ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
PRECIO	0,24	PRECIO	0,12
CALIDAD	0,23	CALIDAD	0,11
RECEPCIÓN	0,23	RECEPCIÓN	0,18
BAÑO	0,22	BAÑO	0,12
LIMPIEZA	0,22	LIMPIEZA	0,09
ATENCIÓN	0,21	ATENCIÓN	0,01
DUDA	0,21	DUDA	0,05
ZONA	0,21	ZONA	0,09
PARKING	0,17	PARKING	0
INSTALACIONES	0,17	INSTALACIONES	0,06
TRATO	0,17	TRATO	0
SPA	0,15	SPA	0,05
COVID	0,13	COVID	0
PRECOVID	0,01	PRECOVID	0
LUJO	0,12	LUJO	0,01
DETALLE	0,06	DETALLE	0,01
DETALLES	0,10	DETALLES	0,01
CAMA	0,34	CAMA	0,17
DECORACIÓN	0,08	DECORACIÓN	0,03
ESTADO	0,09	ESTADO	0,04

Lo primero que salta a la vista en la Tabla 62 es la palabra «precio», cuya frecuencia en el corpus de reseñas escritas originalmente en español (0,24 %) es bastante mayor que su frecuencia en el corpus de reseñas traducidas con TA (0,12 %). Esto es muy significativo, ya que muestra como los hablantes nativos de español de España, miran más el precio que los hablantes nativos de inglés a la hora de decidirse por un sitio u otro en el que alojarse. De manera similar, la palabra «calidad» se ha utilizado mucho más en el corpus de reseñas escritas en español (0,23 %) que en las reseñas traducidas con TA (0,11 %). Lo mismo pasa con «recepción» (0,23 % en el corpus monolingüe de reseñas en español y 0,18 % en el corpus de reseñas traducidas con TA) y «atención» (0,21 % en el corpus monolingüe de reseñas en español y 0,01 % en el corpus de reseñas traducidas con TA).

Otro dato que resulta interesante es que la palabra «limpieza» tiene un elevado porcentaje de ocurrencias (0,22 %) en el corpus monolingüe de reseñas escritas en español, mientras que en el corpus de reseñas traducidas con TA este porcentaje es mucho menor (solo del 0,08 %). En este caso, esta diferencia se podría atribuir al hecho de que, en inglés, la expresión para decir que algo estaba limpio o que se valora la limpieza del hotel, se hace a través expresiones con adjetivos como *it was clean*, y no tanto utilizando el sustantivo correspondiente *cleanliness*. En

español, por otro lado, es tan frecuente el uso del adjetivo «limpio/limpia» como su correspondiente sustantivo «limpieza», por lo que no es de extrañar que ambas palabras sean más frecuentes en el corpus de reseñas escritas originalmente en español. En el corpus de reseñas escritas en español, también aparece la palabra «COVID», refiriéndose al virus que causó la pandemia en 2020, en un 0,13 % de ocasiones. Sin embargo, este no se menciona en el corpus de reseñas traducidas al español mediante TA. Algo similar para con la palabra «PRECOVID», que, aunque en el corpus de español solo aparece en un 0,01 %, en el corpus de reseñas traducidas con TA, tampoco se menciona.

Además, en el corpus monolingüe de reseñas de hoteles escritas en español, también se valoran más las instalaciones, ya que esta misma palabra aparece en un 0,17 % de ocasiones en dicho corpus, mientras que, en el corpus bilingüe de reseñas traducidas con TA, esta solo aparece en un 0,06 % de ocasiones. De entre las instalaciones que más valoran los españoles se han encontrado las palabras «spa» (0,15 % en el corpus de reseñas en español y 0,05 % en el corpus de reseñas traducidas con TA) y «parking». Si se analiza con más detalle la palabra «parking», se observa que esta cuenta con un ,17 % de ocurrencias en el corpus de reseñas escritas en español, mientras que no se han producido ocurrencias en el corpus de reseñas traducidas con TA. Esto se podría deber a que los turistas de otros países no suelen viajar al extranjero en coche, sino en avión, algo que ya se pudo observar anteriormente gracias a la gran ocurrencia en el corpus de reseñas traducidas con TA de palabras como «aeropuerto» y otras palabras relacionadas con el transporte público. Así, al no disponer de coche propio durante el viaje, dichos turistas no suelen preocuparse tanto de si el hotel cuenta con *parking* o no.

Por último, los hablantes nativos de español también parecen valorar aspectos de la apariencia del hotel como la decoración, el lujo, los detalles o el estado del hotel. Así, la palabra «decoración» se repite en un porcentaje del 0,08 % en el corpus de reseñas en español, y solo un 0,03 % en el corpus de reseñas traducidas con TA; la palabra «lujo» aparece un 0,12 % en el corpus de reseñas en español y solo un 0,01 % en las reseñas traducidas con TA; la palabra «detalles» tiene una ocurrencia del 0,10 % en el corpus de reseñas en español y solo 0,01 % en el corpus de reseñas traducidas con TA; y la palabra «estado» cuenta con un porcentaje de ocurrencias del 0,09 % en el monolingüe de reseñas en español y 0,04 % en el bilingüe de reseñas traducidas con TA.

En la Tabla 63, a continuación, se muestran los datos referentes a los adjetivos más utilizados en el corpus de reseñas escritas originalmente en español:

Tabla 63. Adjetivos más frecuentes en reseñas de hoteles corpus monolingüe.

Frecuencia ES Nativo		Frecuencia en ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
BUENA	0,36	BUENA	0,22
BUENAS	0,06	BUENAS	0,04
LIMPIO	0,20	LIMPIO	0,14
MEJOR	0,19	MEJOR	0,10
GRANDE	0,16	GRANDE	0,10

Destaca la menor variedad de adjetivos utilizados en las reseñas escritas en español por hablantes nativos de esta lengua, frente a la gran variedad de adjetivos utilizados en el corpus bilingüe de reseñas traducidas al español utilizando TA. En este caso, se observa como los hablantes de español, utilizan de manera mucho más frecuente algunos adjetivos como «buena» (0,36 % de ocurrencias en el corpus de reseñas en español y 0,22 % en el corpus de reseñas traducidas con TA), «mejor» (0,19 % de ocurrencias en el corpus de reseñas en español y 0,10 % en el corpus de reseñas traducidas con TA) o «grande» (0,16 % de ocurrencias en el corpus de reseñas en español y 0,10 % en el corpus de reseñas traducidas con TA).

### 3.5.2.3. Datos sobre palabras con frecuencia similar en ambos corpus. Corpus hoteles

La última parte de este análisis se ha centrado en examinar qué palabras han aparecido con una frecuencia similar en el corpus de reseñas de hoteles traducidas con TA y en el corpus de reseñas escritas originalmente en español. De esta manera, se han observado hasta qué punto se asemejan las reseñas traducidas con TA y las reseñas escritas originalmente en español. Los resultados de las palabras con frecuencias similares se detallan a continuación (Tabla 64).

Tabla 64. Palabras con frecuencia similar en ambos corpus. Corpus hoteles.

Frecuencia ES Nativo		Frecuencia en ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
ESTANCIA	0.15	ESTANCIA	0.11
ESTRELLAS	0.15	ESTRELLAS	0.16
MAR	0.13	MAR	0.12
VISTAS	0.10	VISTAS	0.06
DUCHA	0.10	DUCHA	0.08
BUFFET	0.09	BUFFET	0.10
CENTRO	0.23	CENTRO	0.20
TERRAZA	0.06	TERRAZA	0.07
EUROS	0.03	EUROS	0.06
SEMANA	0.09	SEMANA	0.07
NIÑOS	0.03	NIÑOS	0.06

Los resultados de la Tabla 64 anterior muestran cómo ambas culturas valoran de manera similar sus estancias utilizando palabras como «estancia», «estrellas» o «semana». De la palabra



«estancia» se han encontrado un 0,15 % de ocurrencias en el corpus de reseñas escritas en español y un 0,11 % en el corpus de reseñas traducidas con TA. Con la palabra «estrellas» pasa algo similar, ya que esta aparece en un 0,15 % de ocasiones en el corpus de reseñas escritas en español y un 0,16 % en el corpus de reseñas traducidas con TA. Por último, la palabra «semana», que aparece 0,09 % de ocurrencias en el corpus de reseñas escritas en español y un 0,07 % en el corpus de reseñas traducidas con TA, lo que puede deberse en parte a que los usuarios pertenecientes a ambas culturas pasan un aproximadamente el mismo tiempo en su destino de vacaciones.

En cuanto a las valoraciones relacionadas con las cualidades del hotel, los usuarios han evaluado de manera similar las «vistas» desde la habitación (0,10 % de ocurrencias en el corpus de reseñas escritas en español y un 0,06 % en el corpus de reseñas traducidas con TA), la «ducha» en sus cuartos (0,10 % de ocurrencias en el corpus de reseñas escritas en español y un 0,08 % en el corpus de reseñas traducidas con TA), el «buffet» del desayuno del hotel (0,09 % de ocurrencias en el corpus de reseñas escritas en español y un 0,10 % en el corpus de reseñas traducidas con TA) y la «terrazza» del bar del hotel (0,06 % de ocurrencias en el corpus de reseñas escritas en español y un 0,07 % en el corpus de reseñas traducidas con TA).

Otra de las palabras con un porcentaje de ocurrencias similar es «mar» (0,13 % de ocurrencias en el corpus de reseñas escritas en español y un 0,12 % en el corpus de reseñas traducidas con TA), ya que, como Valencia es una ciudad costera, los turistas valoran mucho la cercanía y las vistas al mar del hotel. Asimismo, también han valorado la zona de los hoteles, y su ubicación con respecto al «centro» de la ciudad (0,23 % de ocurrencias en el corpus de reseñas escritas en español y un 0,20 % en el corpus de reseñas traducidas con TA). Como muchas familias viajan con sus hijos pequeños, la palabra «niños» se ha repetido en un 0,03 % en el corpus de reseñas escritas en español y un 0,06 % en el corpus de reseñas traducidas con TA. Por último, resulta curioso que la palabra «euros» tenga un porcentaje tan similar en ambos corpus (0,03 % de ocurrencias en el corpus de reseñas escritas en español y un 0,06 % en el corpus de reseñas traducidas con TA), ya que, en los resultados de las palabras más frecuentes en el corpus de reseñas en español, se observó que los españoles piensan mucho más en el dinero a la hora de planear sus vacaciones que los hablantes nativos de inglés.

Como en los análisis anteriores, a continuación, se presenta una pequeña tabla con los resultados de los adjetivos con frecuencias similares en el corpus de reseñas escritas en español y en el corpus de reseñas traducidas con TA (Tabla 65).

Tabla 65. Adjetivos con frecuencia similar en ambos corpus. Corpus hoteles.

Frecuencia ES Nativo		Frecuencia en ES-TA	
Palabra	% ocurrencias	Palabra	% ocurrencias
BUENO	0.14	BUENO	0.18
BUENOS	0.02	BUENOS	0.06
ÚNICO	0.08	ÚNICO	0.08

Al igual que en los análisis anteriores, se observa que hay muy pocos adjetivos que se utilicen de manera similar en ambos corpus. Por este motivo sería posible concluir que los usuarios hablantes nativos de español y los nativos de inglés, no valoran las cosas de la misma forma. En esta ocasión las únicas similitudes encontradas han sido los adjetivos «bueno» (0,14 % de ocurrencias en el corpus de reseñas escritas en español y un 0,18 % en el corpus de reseñas traducidas con TA), «buenos» (0,02% de ocurrencias en el corpus de reseñas escritas en español y un 0,06 % en el corpus de reseñas traducidas con TA) y «único» (en ambos casos un 0,08 % de ocurrencias).

### 3.6. Observaciones análisis cuantitativo y cualitativo

Los resultados de este estudio son representativos de la situación de la TAN de las reseñas generadas por los usuarios en plataformas digitales de reseñas en línea como TripAdvisor.com. De esta manera, se demostrará la necesidad de llevar a cabo una estrategia de posesición apropiada en que también se localicen los contenidos. En el presente análisis, se han estudiado dos tipos de reseñas, las de hoteles y las de restaurantes, ya que cada una de ellas presenta características distintas. En el caso de los hoteles, las principales problemáticas se han producido porque la TAN no traduce correctamente los nombres de lugares históricos o turísticos de la ciudad de Valencia, así como los nombres de los barrios, las calles u otras denominaciones. En el caso de los restaurantes de comida tradicional valenciana, se ha observado una gran incidencia de nombres regionales, por lo que la TAN ha experimentado grandes dificultades para traducirlos. Así, estos errores se han clasificado según diferentes criterios, gracias a los cuales se han obtenido resultados relevantes sobre las problemáticas que dificultan la localización de las reseñas.

En resumen, los principales errores cometidos por la TAN tanto en el corpus de restaurantes como en el de hoteles han sido los siguientes:

1. Uso equivocado del verbo «tener» cuando en inglés aparece *to have*.
2. Uso equivocado o intercambiado de los verbos «ser» y «estar» cuando en inglés aparece *to be*.

3. Expresiones idiomáticas del inglés traducidas literalmente al español o no traducidas.
4. Uso de palabras de variantes del español de Latinoamérica, cuando en este caso la traducción era al español de España.
5. Uso incorrecto de las oraciones de relativo.
6. Uso incorrecto o abusivo de oraciones pasivas.

En concreto, la TAN ha presentado grandes problemas a la hora de traducir palabras con fuertes cargas culturales como, por ejemplo, los platos o ingredientes típicos de la comida valenciana, así como los lugares emblemáticos de la ciudad, los lugares turísticos y los tipos de hoteles y habitaciones, en el caso de las reseñas de hoteles. Uno de los errores más recurrentes en cuanto a este tipo de palabras es que la TAN no ha utilizado un término correcto ni unificado. Es decir, para una misma palabra, la TAN ha utilizado denominaciones que eran erróneas en español, y ha traducido una misma palabra de maneras diferentes en cada reseña o incluso dentro de una misma reseña. Es el caso de *prawns* que se ha traducido como «gambas», «camarones» o «carabineros», entre otras denominaciones.

Por otro lado, ha resultado interesante observar las causas que afectan a los niveles de calidad de la TAN, como han sido los errores de explicación / duplicidad, los problemas causados por errores en el texto original y los errores a causa de la falta de calidad del motor de TAN. Como se ha observado en apartados anteriores, en el caso de las reseñas de restaurantes los tipos de errores más comunes han sido los de explicación / duplicidad. Esto ha ocurrido por la gran cantidad de alimentos regionales que aparecían escritos en español en la reseña original con sus equivalentes en inglés entre paréntesis que, al traducirse con TAN, se han repetido causando una duplicidad innecesaria. Pasa lo mismo cuando, en vez de un equivalente entre paréntesis, lo que aparece es una explicación de qué es el producto. De haberse implementado una estrategia de localización y posesición adecuada, se hubiera concluido que estas palabras equivalentes o las explicaciones son innecesarias para un hablante nativo de español, por lo que se hubieran eliminado. En el caso de los hoteles, como la carga cultural de las reseñas era mucho menor, no se han encontrado tantos casos de este tipo de errores.

Los análisis de ocurrencias han permitido observar diferencias significativas entre los hablantes nativos de inglés y los hablantes nativos de español en cuanto a la forma de valorar los restaurantes y los hoteles. Por ejemplo, en el corpus de restaurantes de reseñas escritas en español, las denominaciones de los alimentos son mucho más precisas y variadas. Así, mientras en las reseñas traducidas con TAN se utiliza la palabra «restaurante» en todas las ocasiones, en el corpus de reseñas escritas en español se utiliza tanto «restaurante» como «arrocería»,

término mucho más acertado y común entre los nativos para denominar este tipo de locales. De la misma forma, cuando se habla de marisco, en inglés suelen aparecer las palabras «gambas», «mejillones», «carabineros» o «camarones». En las reseñas escritas en español, aparece mucha más variedad de denominaciones, por lo que se han aparecido palabras como «gambas», «mejillones», «carabineros», «bogavante», «clóchina» o «gambón», entre otros. Es decir, nombres mucho más precisos y locales para denominar cada tipo marisco. Además, en las reseñas de restaurantes traducidas con TAN, se valoran más a menudo palabras como «tapa» o «tapas», mientras que en las reseñas de español esto no ocurre, ya que se trata de arrocías donde no es tan común que los locales consuman tapas.

En las reseñas en inglés, también se ha apreciado una mayor variedad de adjetivos calificativos positivos tanto en el caso del corpus de reseñas de hoteles como en el de las de restaurantes. En español, las reseñas son más concisas y con menos adjetivos (cuesta más dar valoraciones extremadamente positivas), y se utilizan sobre todo adjetivos como «rico» o «bueno», mientras que en inglés además de estos, se utilizan otros muchos adjetivos con un grado mayor de positividad como «excelente» o «increíble». En esta línea, en inglés ha aparecido en un gran número de ocasiones la expresión «era una joya» para describir los restaurantes. Esta provenía de la expresión *it was a gem*, pero en español no es una expresión común, por lo que cuando la TAN la ha traducido literalmente, el resultado ha sido una traducción poco idiomática.

Otro aspecto interesante es la manera en que dos culturas diferentes expresan un mismo concepto. En muchos casos lo que en inglés se expresa con un solo adjetivo, en español sería necesario reformular o hacer una oración completa. Es el caso de adjetivos como *helpful* en inglés, o la palabra «limpieza» en español. En cuanto al adjetivo *helpful*, la TAN lo ha traducido en todas las ocasiones como «servicial», un adjetivo que denota un cierto grado de inferioridad en español cuando nos referimos al personal de un hotel. En este caso, la mejor opción para expresar el significado de *helpful* en este contexto tal vez habría sido reformular y utilizar una expresión como «nos ayudaron mucho» o alguna expresión similar. De la misma forma, el equivalente en inglés del sustantivo «limpieza» (*cleanliness*) no es tan común, ya que en inglés son mucho más frecuentes expresiones como *it was clean*, («estaba limpio»), es decir, se utiliza mucho más el adjetivo *clean* que su correspondiente sustantivo, por lo que la palabra «limpieza» aparece en muchas menos ocasiones en el corpus de reseñas traducidas con TAN.

Por último, se ha observado con claridad que el uso de palabras relacionadas con el transporte público ha sido mucho menor en el corpus de reseñas escritas originalmente en español. La frecuencia de estas palabras es mucho mayor en el caso de las reseñas traducidas con TAN,

sobre todo por la aparición de palabras como «aeropuerto» en dicho corpus. Asimismo, palabras relacionadas con el transporte privado, como «coche» o «aparcamiento», aparecen mucho más en el corpus de reseñas escritas originalmente en español, ya que los turistas de otros países no se mueven en coche por España tan a menudo como los propios españoles.

Este análisis ha permitido apreciar claramente que la TAN todavía no tiene en cuenta muchos aspectos cruciales a la hora de traducir reseñas con cargas culturales altas. Estos aspectos han sido de tipo lingüístico, ya que la TAN todavía comete errores, que serían evitables con una correcta posesición, pero también de tipo cultural. Son precisamente estos últimos los que pueden causar los mayores malentendidos y delatar cuando una reseña no ha sido escrita originalmente en la lengua del *locale* meta. Con todo esto, se ha podido concluir que, aunque la TAN pueda resultar útil para que el usuario se forme una idea general de lo que puede encontrarse en cierto hotel o restaurante, resultaría muy beneficioso para dichos negocios poseer los contenidos siguiendo los principios de la localización.

### 3.7. Análisis de relación entre la longitud de las reseñas y la calidad del resultado de la TA

El siguiente análisis se ha llevado a cabo para determinar si la longitud de las reseñas afecta a la calidad de los resultados de la traducción automática. Los principales autores en materia de contenido generado por el usuario, como Vásquez (2016) no hacen referencia a la longitud de las reseñas como factor determinante de la calidad del contenido en sus investigaciones. Por este motivo, para facilitar el estudio de cómo afecta la longitud de las reseñas a la calidad de los resultados de TAN, en esta tesis se ha elaborado una clasificación propia ad hoc según los datos del corpus analizado, en que se han categorizado las reseñas en reseñas cortas y reseñas largas, según el número de palabras en cada reseña.

La bibliografía consultada (Castilho et al., 2018; Cho et al., 2014; Fiederer y O'Brien, 2009; Pouget-Abadie et al., 2014) establece una media de 20 palabras como extensión típica de las oraciones, considerando largas las oraciones de unas 30 palabras y cortas las oraciones de unas 10 o menos. Estos estudios se han aplicado a contextos de cursos online, noticias, transcripciones de conferencias, informática y software, y procedimientos del Parlamento Europeo, entre otros. Sin embargo, no se han encontrado estudios que analizaran cómo afecta la longitud a la calidad de las reseñas de hoteles y restaurantes escritas en portales de turismo online y traducidas con TAN.

Así, se ha seguido la premisa de que cuanto más largas son las reseñas, mayor es la complejidad sintáctica y, por tanto, peor será la calidad de los resultados de TAN. Para ello, primero se ha calculado la longitud media de las reseñas de cada corpus. En el caso de las reseñas de hoteles, la longitud media de las reseñas ha sido de 110 palabras; en el caso de las reseñas de restaurantes, la longitud media ha sido menor, con 70 palabras por reseña. Después se han calculado los porcentajes correspondientes a la cantidad de reseñas largas y de reseñas cortas. Finalmente, se ha calculado el porcentaje de reseñas cortas con calidad nativa, correcta, aceptable o inaceptable y, posteriormente, se ha hecho lo mismo con las reseñas largas. En los siguientes apartados, se detallan los resultados del análisis cuantitativo sobre la longitud de las reseñas, tanto en el corpus de reseñas de hoteles como en el de reseñas de restaurantes.

La Tabla 66 (a continuación) muestra los datos generales sobre el número de reseñas en cada corpus, así como la media de oraciones en estos, el número de frases largas y el número de frases cortas.

Tabla 66. Datos generales sobre reseñas.

	Reseñas corpus	en	Núm. De frases	Núm. De frases largas	Núm. De frases cortas
<b>Hoteles</b>	250		1736	354	1381
<b>Restaurantes</b>	250		1342	237	1104

A modo de ejemplo en la Tabla 67, a continuación, se muestran algunos ejemplos representativos de las diferentes longitudes de reseñas encontradas:

Tabla 67. Ejemplos de reseñas y longitudes.

	TO	TA_ES
<b>Ejemplo 1</b> <b>Frases largas (+20 palabras)</b>	<i>finally as we were there so long I think we had dealings with every front of house staff that worked there and except one lady , everyone else were happy, smiling , polite and a joy to be served by... but the was one lady who was rude , refused to look at you when talking to you , took your plates before you had finished , never smiled ... you</i>	Finalmente, mientras estuvimos allí tanto tiempo, creo que tuvimos tratos con todos los miembros del personal de la casa que trabajaban allí y, excepto una señora, todos los demás estaban felices, sonrientes, educados y felices de ser atendidos ... pero fue una señora que fue grosero, se negó a mirarte cuando te hablaba,

---

<p><b>Ejemplo 2</b> Frase promedio (aproximadamente 20 palabras)</p>	<p><i>must all know who I am talking about ??... we ended up avoiding the bar and restaurant when she was on duty ... she had no place working in such a friendly great hotel ... this person does not fit in at all to your lovely hotel</i></p>	<p>tomó tus platos antes de que hubieras terminado, nunca sonrió ... ¡¡¡itodos deben saber de quién estoy hablando ¿? ... terminamos evitando el bar y el restaurante cuando estaba de servicio ... no tenía lugar para trabajar en un hotel tan agradable y agradable ... esta persona no encaja en absoluto en su encantador hotel</p>
<p><b>Ejemplo 3</b> Frase corta (-20 palabras)</p>	<p><i>We ate basically nothing and then found another breakfast room which appeared to be chaotic with queues out the door.</i></p> <p><i>Highly recommended</i></p>	<p>Básicamente no comimos nada y luego encontramos otra sala de desayuno que parecía ser caótica con colas por la puerta.</p> <p>¡Muy recomendable!</p>

---

Como se puede observar en los ejemplos de frases muy cortas, la TAN es prácticamente imperceptible, sin embargo, en el ejemplo de reseña larga, a medida que la longitud aumenta, la presencia de la TAN se hace mucho más patente, ya que la calidad empeora.

### 3.7.1. Parámetros de calidad utilizados

Evaluar los sistemas de TAN es esencial para que estos se puedan mejorar, sin embargo, todavía no hay una definición consensuada de calidad de TAN entre el mundo académico y la industria de la traducción y la localización (Castilho et al., 2017: 109). No obstante, en los últimos años, se ha intentado unificar el proceso de evaluación gracias a los estándares de calidad ISO/TS 11669:2012 y ISO 18587:2017, y a algunas asociaciones como la desaparecida LISA o como TAUS. Según Lommel, Uszkoreit y Burchardt (2014: 455) uno de los principales problemas de estas métricas de calidad es la falta de estandarización. Doherty (2017: 132), por su parte, atribuye la falta de acuerdo con que algunas de las métricas para evaluar la calidad se basan en el texto original y otras en el texto meta. Además, los diferentes tipos de texto y pares de lenguas, así como la complejidad del proceso de traducción, dificultan la fijación de un objetivo de calidad unificado (Doherty, 2017: 132).

El nivel de calidad de la TAN se puede medir a través de mecanismos automáticos de evaluación, como BLEU, o con métricas manuales como MQM (Multidimensional Quality Metrics), desarrolladas gracias al proyecto QTLaunchPad, financiado por la Unión Europea, o el Dynamic Quality Framework (DQF) de TAUS, que es el modelo más utilizado en la actualidad (Lommel, Uszkoreit y Burchardt, 2014: 461).

Según Valli (2015: 127), el modelo dinámico de TAUS gira en torno a tres parámetros de evaluación (*utility*, *time* y *sentiment*) y, cada uno de ellos contará con mayor o menor relevancia según el contenido que se vaya a traducir. TAUS define «utilidad» (*utility*) como la funcionalidad del contenido, «tiempo» (*time*) como la velocidad a la que se llevó a cabo la traducción y «sentimiento» (*sentiment*) como la forma en que los usuarios perciben la imagen de cierta marca (O'Brien et al., 2011). Este modelo evalúa algunos parámetros como la precisión (*accuracy*), la fluidez (*fluency*), la terminología (*terminology*), el estilo (*style*), el diseño (*design*), las convenciones de los *locales* (*locale conventions*), o la veracidad (*verity*). Además, el modelo DQF cuenta con varios niveles de gravedad para las diferentes categorías de error (Valli, 2015).

Como ya se describió en la sección de metodología de esta tesis, se ha elaborado una clasificación propia cuyos principales criterios de calidad se han basado en una adaptación del modelo DQF de TAUS, pero que, en este caso, ha puesto especial énfasis en las convenciones propias del *locale* meta, ya que uno de los objetivos era comprobar si la TAN cumple con los principios de la localización.

De esta manera, se han identificado errores lingüísticos y culturales causados por la calidad del motor de TAN, por errores de duplicidad y explicación, así como por errores en el texto original. Relacionado con este último tipo de error, otro factor que ha podido afectar a la calidad ha sido la aparición de omisiones o errores en el uso de signos de puntuación. Por ejemplo, se han omitido puntos o se han utilizado comas en exceso, ya que al tratarse de un género con rasgos de oralidad y escrito de manera muy informal, no siempre se han seguido las reglas de ortografía y no se corrigen los textos de manera exhaustiva antes de publicarlos. Esto además se agrava cuanto más largo es el texto. Algunos ejemplos se muestran a continuación (Tabla 68):

Tabla 68. Ejemplos de errores por omisiones o uso incorrecto de signos de puntuación.

TO	ESTA	Comentario
<i>Either they have this type of guest often, so they are blinded to how illogical their conclusions</i>	O tienen este tipo de invitado con frecuencia, por lo que no saben qué tan ilógicas son sus	En este sentido la TAN ha cambiado por completo la oración.



---

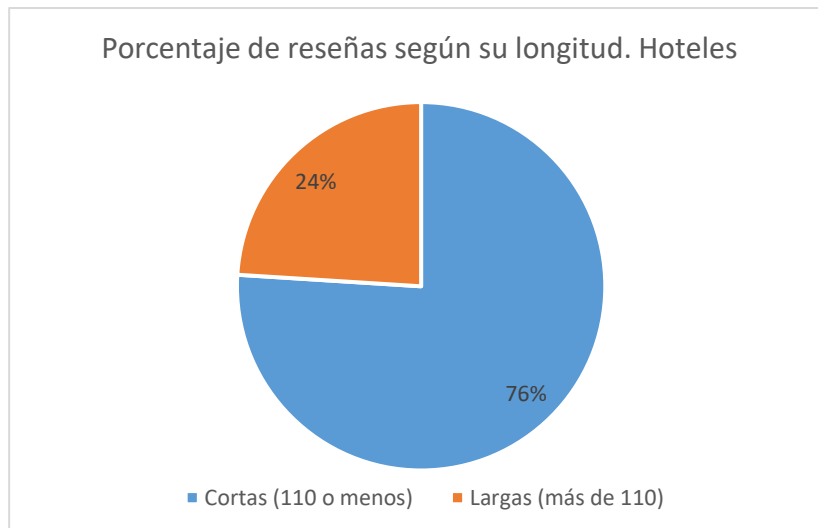
<p><i>are on all guests—<b>or they don't</b> have enough experience in hospitality [...]</i></p>	<p>conclusiones para todos los invitados, <b>o no lo hacen</b>.</p>	
<p><i>Friendly.. my boyfriend proposed to me there and owner helped him to 208ill208ize all.. I 208ill go back, back and back..</i></p>	<p>El servicio del personal fue realmente amable ... mi novio me propuso matrimonio allí y el propietario lo ayudó a organizar todo ... Volveré, volveré y volveré ...</p>	<p>En este caso en el original se han utilizado de manera incorrecta los puntos suspensivos y la TAN, a su vez, los ha utilizado incorrectamente en español, ya que ha añadido un espacio antes de los puntos suspensivos.</p>
<p><i>We have a great waiter <b>Saul</b>We will be <b>backGreat</b> Paella Chicken and <b>rabbitGreat</b> <b>SeafoodVery</b> generous portion of meat and <b>seafoodBest</b> <b>tiramisuOnce</b> again thank you</i></p>	<p>tenemos un gran <b>camareroSaul</b> <b>VolveremosGran</b> paella Pollo y <b>conejoGran</b> <b>mariscoMuy</b> generosa porción de carne y <b>mariscosMejor tiramisúUna</b> vez más, gracias</p>	<p>En este caso en el original hubo ejemplos de palabras que no se separaron correctamente con un espacio. La TAN en vez de corregir el error, lo ha conservado en la traducción al español y ha dado como resultado un texto complicado de seguir.</p>

---

Este tipo de errores influye en la manera en que el motor de TAN realiza la traducción, puesto que procesa como oraciones completas, algunas oraciones truncadas o mal separadas y da como resultado traducciones sin sentido, lo que se agrava cuanto más larga es la reseña, ya que el número de errores de este tipo es mayor. En el apartado siguiente, se presentan los resultados del análisis de cómo la longitud afecta a la calidad de los textos traducidos con TAN.

### 3.7.2. Calidad de las reseñas de hoteles según su longitud

En el caso de los hoteles, la longitud media de estas reseñas es de 110 palabras por reseña. En la Ilustración 30, a continuación, se detallan los porcentajes de reseñas cortas (110 palabras o menos) y de reseñas largas (más de 110 palabras) encontrados.



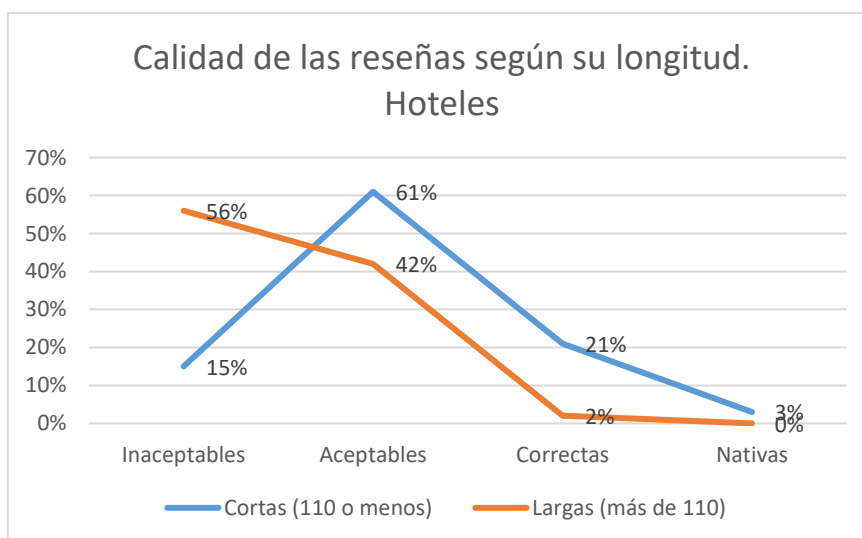
*Ilustración 30. Porcentajes de reseñas según su longitud. Corpus hoteles.*

Del total de reseñas de hoteles, un 76 % tenía 110 palabras o menos, por lo que se han considerado como reseñas cortas, y solo un 24 % tenía más de 110 palabras, por lo que se han considerado como reseñas largas.

Para determinar si la longitud de las reseñas ha afectado a la calidad, se ha analizado el porcentaje de reseñas cortas y largas con nivel de calidad nativo, correcto, aceptable e inaceptable, los resultados se muestran en la Tabla 69 y la Ilustración 31 a continuación:

*Tabla 69. Porcentajes de reseñas según su calidad y longitud. Corpus hoteles.*

	Inaceptables	Aceptables	Correctas	Nativas
<b>Cortas (110 o menos)</b>	15%	61%	21%	3%
<b>Largas (más de 110)</b>	56%	42%	2%	0%



*Ilustración 31. Porcentajes de reseñas según su calidad y longitud. Corpus hoteles.*

Como se observa en la Tabla 69 y en la Ilustración 31, las reseñas cortas cuentan en su mayoría con una calidad aceptable (61 %), seguidas de las reseñas con calidad correcta (21 %), las inaceptables (15 %) y las nativas (3 %). En el caso de las reseñas largas, más de la mitad cuenta con un nivel de calidad inaceptable (56 %), seguido por el nivel de calidad aceptable (42 %), y por el nivel de calidad correcto (2 %), pero sin ninguna reseña con nivel de calidad nativo.

Las principales diferencias entre las reseñas largas y las cortas son los porcentajes del nivel de calidad inaceptable y del nivel nativo. Por un lado, destaca la gran diferencia entre el porcentaje de reseñas cortas con un nivel de calidad inaceptable (15 %) y el de reseñas largas con este mismo nivel de calidad (56 %), ya que el porcentaje es mucho mayor en el caso de las reseñas largas que en de las cortas. Por otro lado, también es significativo que no haya ni una sola reseña larga con nivel de calidad nativo, pero que sí que haya algunas reseñas cortas que hayan alcanzado este nivel (3%).

### 3.7.3. Calidad de las reseñas de restaurantes según su longitud

En este apartado, se presentan los datos correspondientes al corpus de reseñas de restaurantes. En este caso, la longitud media de estas reseñas es de 70 palabras por reseña. A continuación, la Ilustración 32 muestra los porcentajes de reseñas cortas (70 palabras o menos) y de reseñas largas (más de 70 palabras) de este corpus.

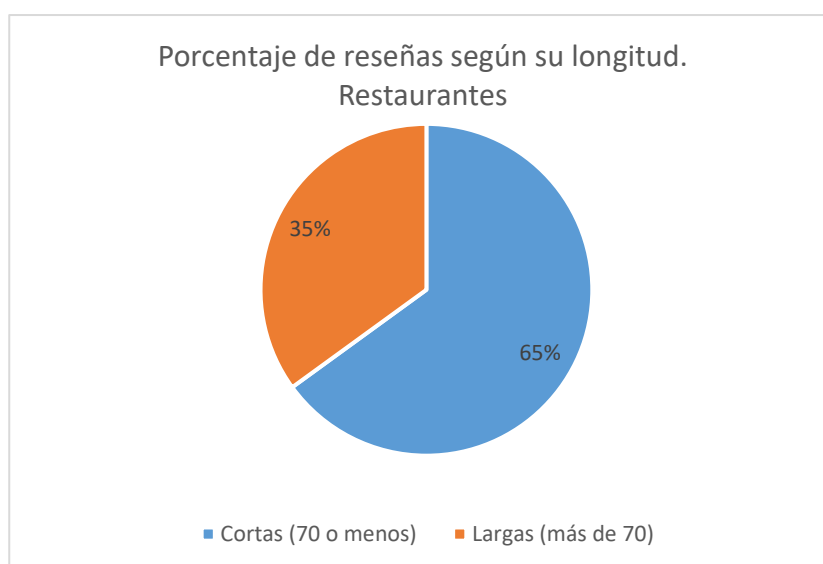


Ilustración 32. Porcentajes de reseñas según su longitud. Corpus restaurantes.

Del total de reseñas de restaurantes, un 65 % tenía 70 palabras o menos, es decir, eran reseñas cortas, y solo el 35 % tenía más de 70 palabras, por lo que se han considerado como reseñas largas. De esta manera se observa que las reseñas de restaurantes tienden a ser bastante

concisas. Los resultados del análisis de los porcentajes de reseñas cortas y largas y sus respectivos niveles de calidad se muestran en la Tabla 70 y la Ilustración 33, a continuación:

Tabla 70. Porcentajes de reseñas según su calidad y longitud. Corpus restaurantes.

	Inaceptables	Aceptables	Correctas	Nativas
<b>Cortas (70 o menos)</b>	13%	58%	24%	5%
<b>Largas (más de 70)</b>	16%	70%	14%	0%

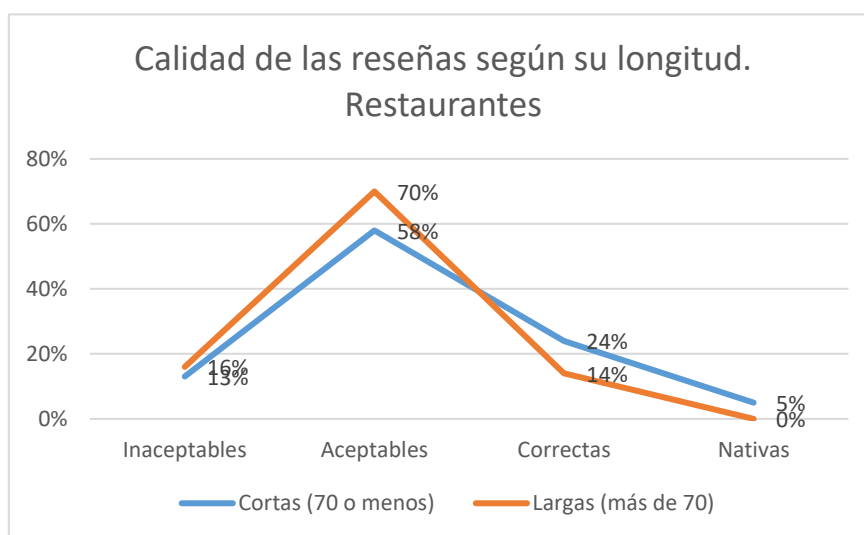


Ilustración 33. Porcentajes de reseñas según su calidad y longitud. Corpus restaurantes.

La Tabla 70, muestra los porcentajes de reseñas cortas y de reseñas largas junto con sus niveles de calidad. De esta manera se observa que, en el caso de los restaurantes, más de la mitad de las reseñas cortas tienen un nivel de calidad aceptable (58 %). Las reseñas cortas con nivel correcto son las segundas con mayor porcentaje (24 %), seguidas de las inaceptables (13 %) y las nativas (5 %). De igual forma, las reseñas largas, son en su mayoría aceptables (70 %). En este caso el porcentaje de reseñas inaceptables (16 %) y correctas (14 %) es similar, aunque el porcentaje de reseñas inaceptables sea algo mayor. De nuevo, en el caso de las reseñas largas, no hay ni una sola con nivel nativo.

Al igual que en el caso de las reseñas de hoteles, una de las mayores diferencias entre las reseñas largas y las cortas en cuanto a los niveles de calidad, ha sido que, mientras que en las reseñas cortas sí se han encontrado casos de reseñas nativas, en el caso de las reseñas largas no se ha encontrado ningún caso de reseña con nivel de calidad nativo. En este caso también se observa una diferencia considerable entre las reseñas cortas y las reseñas largas con nivel correcto, en que las cortas tienen un porcentaje mayor. No obstante, en las reseñas de restaurantes, se han advertido unos porcentajes similares tanto en las reseñas cortas como en las reseñas largas con

nivel inaceptable, tal vez porque la cantidad de nombres regionales es similar en ambas y en ambos casos se han encontrado traducciones incorrectas.

### 3.8. Observaciones análisis de la longitud de la reseña y calidad

Tras haber analizado ambos corpus, el primer dato que salta a la vista es que las reseñas de restaurantes son mucho más concisas de media que las de hoteles, ya que las de restaurantes tienen una longitud media de 70 palabras por reseña y las de hoteles suelen tener 110 palabras de media por reseña. Esto se puede deber a que las reseñas de hoteles suelen tener un estilo mucho más narrativo en que se suelen contar los momentos previos a la llegada al hotel (viaje en transporte público, cercanía a las estaciones, etc.), la llegada (aparcamiento, *check-in* y primeras impresiones del hotel y las habitaciones), la estancia en sí (monumentos visitados, cercanía de estos al hotel, restaurantes, acontecimientos sucedidos durante la estancia, etc.) y salida (*check-out*, viaje de vuelta a casa, etc.). En el caso de los restaurantes, los usuarios se centran más en describir el local y la comida en sí, por lo que las reseñas suelen ser mucho más concisas. Además, al ser más largas, las reseñas de hoteles tienen un porcentaje más elevado de reseñas incorrectas que las de restaurantes (Tabla 71, a continuación):

Tabla 71. Porcentajes de reseñas inaceptables de restaurantes vs. Hoteles.

	Corpus restaurantes	Corpus hoteles
<b>Cortas inaceptables</b>	13 %	15 %
<b>Largas inaceptables</b>	16 %	56 %

Es más, se ha observado que el porcentaje de reseñas cortas nativas es mayor en el caso de las reseñas de restaurantes que en las de hoteles, aunque no haya habido ninguna reseña larga nativa en ninguno de los dos corpus. Así, también sería posible afirmar que, como las reseñas de restaurantes suelen ser más cortas, el porcentaje de reseñas con calidad nativa es mayor. Esto se detalla en la Tabla 72, a continuación:

Tabla 72. Porcentajes de reseñas nativas de restaurantes vs. Hoteles.

	Corpus restaurantes	Corpus hoteles
<b>Cortas nativas</b>	5 %	3 %
<b>Largas nativas</b>	0 %	0 %

Por otra parte, en ambos corpus (hoteles y restaurantes), el nivel de calidad aceptable ha sido uno de los más frecuentes tanto en el caso de las reseñas largas como en el de las cortas. Tanto en el corpus de restaurantes como en el de hoteles, las reseñas cortas han tenido

predominantemente un nivel de calidad aceptable, seguido de un nivel de calidad correcto, después inaceptable y, por último, nativo. En el caso de las reseñas de hoteles largas, el mayor porcentaje es el de reseñas inaceptables, seguidas de las aceptables y las correctas. No obstante, no ha habido ninguna reseña larga que se pudiera considerar nativa. Por último, en el caso de las reseñas largas de restaurantes, las reseñas aceptables han sido las más frecuentes, seguidas de las inaceptables, las correctas y, de nuevo, ninguna reseña con nivel de calidad nativo. Así, se podría concluir que el que en ambos corpus haya unos porcentajes altos de reseñas largas inaceptables (segundo más algo en el caso de los restaurantes y más alto en el caso de los hoteles) es una muestra de cómo la longitud de la reseña influye en la calidad de los resultados de TA.

En el apartado siguiente se detallan las principales conclusiones extraídas de esta tesis y los análisis llevados a cabo, para determinar si se han cumplido las premisas en las que se ha basado este trabajo. Además, se han propuesto futuras líneas de estudio que podrían surgir a partir de esta tesis.

## 4. Conclusiones

En esta investigación se ha trabajado para comprobar si la calidad de los resultados de TAN de las reseñas generadas por los usuarios en plataformas de turismo de internet se puede evaluar siguiendo los principios de la localización, junto con las métricas existentes sobre calidad de TA. De esta manera, se ha querido comprobar la calidad de los resultados de TAN de los contenidos generados por los usuarios siguiendo una escala basada en las métricas de calidad de TAUS. Tras concluir el apartado de análisis y resultados, esta sección presenta las principales contribuciones de este estudio y propone algunas posibles líneas de investigación futuras en el campo de la TA, de las métricas de evaluación de la calidad y de las reseñas generadas por los usuarios.

Esta investigación ha estado motivada por la creciente cantidad de contenido generado por el usuario que se ha originado gracias al continuo desarrollo de Internet. Entre estos contenidos se encuentra el objeto de estudio de esta tesis, las reseñas generadas por el usuario de hoteles y restaurantes publicadas en plataformas de turismo en línea como TripAdvisor.com. La gran velocidad a la que se genera este tipo de contenidos provoca que se necesiten traducciones de manera más inmediata, por lo que, cada vez más, estos contenidos se traducen utilizando TAN. No obstante, pese a que la TAN produce resultados cada vez mejores, todavía no es perfecta. Por eso, en este estudio se ha dado respuesta a la pregunta de si la TAN produce contenidos de suficiente calidad que sigan los principios de la Localización o si todavía es necesario implementar una estrategia de PE posterior. Además, otro parámetro de calidad que se ha estudiado ha sido la longitud de las reseñas, lo que ha ayudado a comprobar en qué medida esta afecta a la calidad de los resultados de TAN.

Así, en esta tesis se ha trabajado con diferentes premisas. La primera de ellas, que el CGU afecta a la percepción de los usuarios del producto o servicio en cuestión. La segunda premisa con la que se ha trabajado ha sido que, si no se traduce el CGU, ciertos *locales* no podrán recibir esta información, lo que podría terminar afectando a la venta del producto o servicio. La tercera premisa de trabajo ha sido que, si estos contenidos se traducen solo con TAN sin poseer, podrían producirse muchos errores, entre ellos errores de localización, los cuales pueden afectar negativamente a la percepción del producto o servicio turístico por parte de los usuarios. La cuarta premisa ha sido que, aunque la TAN representa una ayuda en este contexto, sigue siendo necesario que un poseedor humano revise los contenidos teniendo en cuenta los principios de la localización. Por último, la quinta premisa en la que se ha basado este estudio ha sido que la longitud de las reseñas afecta a la calidad, por lo que, cuanto más larga sea la reseña, peor será la calidad de los resultados de TAN.

Para realizar los estudios necesarios para comprobar estas premisas, primero se han investigado los antecedentes del CGU, que se origina a partir del WOM y, posteriormente del eWOM, y que crece de manera exponencial con la aparición y la expansión de internet y la Web 2.0. Además, también se han expuesto los diferentes tipos de CGU, entre los que se encuentran las reseñas generadas por los usuarios en plataformas en línea de turismo, y se han explicado los diferentes puntos de vista desde los que se han estudiado estos contenidos.

También se ha detallado la forma en que los usuarios pasan de ser meros consumidores de contenidos a prosumidores, que leen y también crean contenidos. Dichos prosumidores se ven muy influidos por las opiniones de otros usuarios y por la calidad del contenido, y tienden a desconfiar si la calidad de los contenidos no es aceptable, por eso es imprescindible que estos contenidos se gestionen de manera correcta. En el caso de las RGU traducidas con TAN, la correcta gestión del contenido implica la adecuada implementación de una estrategia de posesición por parte de traductores profesionales, de manera que se tengan en cuenta los principios de la localización.

Asimismo, se ha realizado un estudio bibliográfico de la historia de la TA, desde sus inicios aproximadamente en los años 30, hasta la actualidad. Se ha descrito el funcionamiento de los diferentes tipos de TA que han surgido a lo largo de la historia, como la TA basada en reglas, la TA basada en ejemplos, la TAE o la TAN. Además, se ha detallado la arquitectura codificador-decodificador, base de la TAN, así como los principales avances en materia de TA como puede ser el uso de RNN, de CNN, los mecanismos de atención o el modelo *Transformer*. Adicionalmente, se han presentado algunas de las principales métricas de evaluación de la calidad de la traducción automática, haciendo hincapié en el modelo dinámico de TAUS (DQF), el más actual y el más utilizado hoy en día.

Tras esta detallada introducción sobre los aspectos teóricos del CGU, la TAN, la PE y la Localización, se han llevado a cabo una serie de análisis. Por un lado, se han realizado unos análisis cuantitativos y cualitativos, para los que se han seguido parámetros basados en las métricas de calidad de TAUS, con el objetivo de medir la calidad de los resultados de TA, así como para determinar qué errores sigue cometiendo la TA al traducir este tipo de contenidos, de qué gravedad son y si podrían llegar a afectar de manera negativa a la percepción de los consumidores del *locale* meta. Por otro lado, se ha realizado un análisis para determinar en qué medida la longitud de las reseñas afecta a la calidad de los resultados de TAN. A continuación, se detallan las conclusiones extraídas a partir de los resultados obtenidos en los análisis.



Tras llevar a cabo los análisis cuantitativos y cualitativos de los resultados de TAN de las reseñas generadas por los usuarios en la plataforma de turismo TripAdvisor.com, ha sido posible extraer diversas conclusiones sobre la cantidad y los tipos de errores encontrados y sobre cómo afectan estos a la calidad de dichos textos. En general, se han encontrado más errores lingüísticos que culturales, tanto en el corpus de hoteles como en el de restaurantes. Además, prestando atención a los tipos de errores, se ha observado que, en el caso del corpus de reseñas de restaurantes, los errores son sobre todo de tipo explicación/duplicidad, ya que es muy común que aparezcan nombres de alimentos regionales que muchas veces se han explicado o traducido entre paréntesis en el texto original. No obstante, en el corpus de hoteles, esto no pasa tanto porque no aparecen tantos términos regionales. En el caso de los hoteles, los tipos de errores más frecuentes son aquellos producidos por la calidad del propio motor de TAN.

Asimismo, se ha observado que algunos de los principales problemas que surgen al traducir las reseñas con TAN son los problemas de localización culturales y los errores de tipo lingüístico. Por un lado, gracias al análisis del corpus, se han detectado ejemplos de problemas de localización culturales en algunas reseñas (sobre todo de restaurantes, ya que estas contienen un mayor número de términos con alta carga cultural que las reseñas de hoteles), donde la TAN sigue cometiendo muchos errores al traducir nombres de lugares históricos, en el caso de los hoteles, y al traducir nombres de comida tradicional valenciana, en el caso de los restaurantes. Otro problema de localización detectado ha sido que, en muchas ocasiones, la TAN utiliza palabras de variantes latinoamericanas del español. En este estudio, esto se ha considerado erróneo, ya que las reseñas estaban destinadas a un público de España.

Por otro lado, dentro de los errores de tipo lingüístico, se ha observado que la TAN confunde principalmente los verbos «ser/estar» cuando en los textos en inglés aparece el verbo *to be*. Asimismo, la TAN confunde el uso del verbo «tener» cuando en inglés aparece el verbo *to have*, sobre todo en relación con la comida. Además, La TAN traduce al español de manera literal algunas expresiones idiomáticas del inglés. Otro error importante de tipo lingüístico por parte de la TAN ha sido el uso inadecuado y excesivo de oraciones pasivas, así como un uso incorrecto de las oraciones de relativo. También se han detectado muchos ejemplos de traducciones inconsistentes, en que una misma palabra se ha traducido de maneras diferentes. Por ejemplo, la palabra *prawns* que se ha traducido como «carabineros», «gambas» o «camarones» en las diferentes reseñas o, a veces, incluso dentro de una misma reseña. Por lo tanto, sería posible concluir que la TAN todavía comete errores al traducir contenidos con alta carga cultural, como las reseñas de restaurantes u hoteles, principalmente cuando aparecen nombres de lugares y

de alimentos regionales. Además, también sigue teniendo problemas para traducir de manera consistente e idiomática ciertas palabras y expresiones.

Durante el análisis del corpus también se identificaron las principales causas de estos problemas y se detectó que los errores más comunes son aquellos denominados de explicación/duplicidad.

Estos son sobre todo muy comunes en las reseñas de restaurantes, ya que hay muchos ejemplos de alimentos regionales que en las reseñas en inglés aparecen en español con su correspondiente traducción en inglés entre paréntesis, o incluso con una explicación detrás. Al traducirse al español, lo correcto hubiera sido dejar el término traducido y eliminar esa repetición o explicación que no es necesaria para los hablantes nativos del *locale* español. Así, es posible afirmar que, de haberse implementado una estrategia de PE, que siguiera los principios de la localización, por parte de un traductor profesional y nativo del *locale* meta, se hubiera podido evitar este tipo de errores antes de que los usuarios hubieran leído las reseñas traducidas.

En el análisis de ocurrencias también se pudo observar que existen grandes diferencias entre los hablantes nativos de inglés y los de español en cuanto al número de veces en que aparece una misma palabra. Por ejemplo, la palabra «comida» aparece de manera mucho más frecuente en el corpus de inglés que en el de español, al igual que la palabra «tapas». En esta línea, también se ha comprobado que los españoles y los angloparlantes no utilizan adjetivos del mismo grado para realizar sus valoraciones. Así, en inglés, son muy frecuentes palabras como «excelente» o «hermoso», pero en español, es más frecuente el uso de adjetivos con un grado más neutro como pueden ser «bueno/buena». También se han encontrado formas diferentes de expresar un mismo concepto, por ejemplo, cuando en español se refieren a «la limpieza», en inglés no dicen *cleanliness*, sino más bien *it was clean*. Por eso, es posible concluir que los hablantes nativos de inglés y los de español no utilizan el mismo vocabulario en sus valoraciones y, por lo tanto, no valoran de la misma manera los hoteles y los restaurantes. Por eso, si se traduce solo con TAN sin poseer ni localizar, se puede llegar a notar el uso de palabras y expresiones más propias del inglés que no se utilicen tan frecuentemente en español.

En esta tesis también se ha realizado otro análisis para determinar si la longitud de las reseñas ha afectado a la calidad de los resultados de la TAN. A partir los resultados obtenidos se puede concluir que la calidad de los resultados de TA es inferior cuando las reseñas tienen una mayor longitud, especialmente cuando estas contienen altas cargas culturales, como pasa en las reseñas de restaurantes y de hoteles de un *locale* específico. Esto puede deducirse de los resultados que muestran que no hay reseñas largas nativas ni en el corpus de reseñas de

restaurantes ni en el de reseñas de hoteles y que, en ambos corpus, el porcentaje de reseñas cortas inaceptables es menor que el porcentaje de reseñas largas inaceptables. Es más, por el propio estilo más narrativo y detallado de las reseñas de hoteles, estas son por lo general más largas que las de restaurantes, lo que a su vez ha resultado en un número más reducido de reseñas nativas.

A modo de resumen, se podría afirmar que la TAN sigue cometiendo errores que serían evitables si se implementara una correcta estrategia de PE. Estos errores son tanto de tipo lingüístico como de tipo cultural, por lo que es imprescindible que se tengan en cuenta los principios de la localización cuando se posezcan los contenidos. Además, en el caso de los hoteles y los restaurantes, esto es todavía más importante, puesto que los productos y servicios ofrecidos son intangibles y el usuario no los puede probar antes de adquirirlos, por lo que una buena primera impresión es crucial.

Del análisis realizado se ha extraído la conclusión de que la TAN puede ayudar a que los usuarios se formen una idea general de los contenidos. No obstante, esta todavía no es capaz de conseguir traducciones de reseñas con calidad humana. Puesto que el propósito de las reseñas es ayudar a que los usuarios se decidan por un producto o servicio en concreto, implementar una estrategia de posesión que tenga en cuenta los principios de la localización sería muy beneficioso para la imagen de los hoteles y de los restaurantes en cuestión. Si no se hace esto, se podrían perder matices importantes, lo que podría dificultar la comprensión y dar una mala impresión del lugar.

A continuación, se proponen algunas posibles líneas de investigación futuras que podrían surgir a partir de esta tesis.

#### **4.1. Trabajo futuro**

Durante la realización del marco teórico se encontró escasa bibliografía sobre el estudio del CGU en relación con la TAN de este tipo de contenidos. Este estudio se centró en las reseñas de turismo escritas en plataformas de turismo en línea, como TripAdvisor.com, no obstante, existen muchos otros tipos de CGU que también se pueden estudiar, como por ejemplo reseñas de productos, valoración de servicios.

Existen otras posibles líneas de estudio no necesariamente relacionadas con esta tesis, como la realización de un estudio con poseedores profesionales para determinar el esfuerzo de PE de las reseñas escritas en portales de turismo en línea como TripAdvisor.com, a partir de un corpus de reseñas. El esfuerzo de posesión se podría evaluar, por ejemplo, mediante el análisis del

número y el tipo de errores que corrigen los poseedores y el tiempo invertido en su corrección. Otro posible estudio podría ser un análisis de la calidad de las reseñas, extrapolado a las reseñas de otro tipo de productos o servicios vendidos en Internet, como pueden ser las reseñas de productos en Amazon, utilizando también las métricas de calidad de TAUS como guía para evaluar la calidad de dichas reseñas.

Además, hoy en día la TA avanza a gran velocidad, por lo que cada vez existen más modelos de TA. De esta manera, también sería interesante realizar una comparativa de la forma en que los motores de TAN más punteros (como pueden ser DeepL o Google Translate) traducen algún tipo de CGU como podrían ser las reseñas generadas por los usuarios.

Otra posible investigación sería el estudio del *sentiment analysis* relacionado con la TA de las reseñas en línea. Si bien es cierto que ya existen estudios que se basan en *sentiment analysis* de textos traducidos con TA, no existe bibliografía extensa sobre esta temática relacionada con el CGU y las reseñas en línea.

Otra posibilidad sería realizar el mismo estudio llevado a cabo en la presente tesis, pero con otros pares de lenguas o incluso con pares de lenguas minoritarias. Así, se podría observar si la TAN actúa del mismo modo con otras combinaciones de idiomas que con los textos que se traducen del inglés al español.

También sería interesante realizar un análisis en profundidad sobre las características que diferencian un *locale* de otro en cuanto a la manera de expresar un mismo concepto en este tipo de contenidos. En esta tesis se realiza una introducción a este tipo de análisis, pero sería de gran interés llevar a cabo un estudio detallado de estas diferencias. Además, también podría resultar interesante realizar un estudio basado en la premisa planteada en apartados anteriores sobre la longitud de las oraciones de las reseñas y cómo esta afecta a la calidad de los resultados de TA. Es más, dentro de este, se podría realizar un estudio con usuarios reales para evaluar cómo la calidad de la traducción automática de las reseñas afecta a la percepción de estas por parte de los usuarios. Por último, se podría utilizar esta tesis como un apoyo didáctico a la docencia de la posesición en las universidades para formar futuros poseedores, ya que puede ser de gran interés realizar a modo de ejercicio práctico un análisis a pequeña escala de aquellos posibles errores que cometa la TAN al traducir CGU.

## 5. Bibliografía

- Abaitua, J. (2002). La traducción basada en reglas. Recuperado de: [http://paginaspersonales.deusto.es/abaitua/konzeptu/ta/mt10h\\_es/ta10h-4es.htm](http://paginaspersonales.deusto.es/abaitua/konzeptu/ta/mt10h_es/ta10h-4es.htm)
- Al Ansary, S. (2014). Interlingua-based machine translation systems: UNL versus other interlinguas. *The Egyptian Journal of Language Engineering*, 1(1), 42-54.
- Allen, J. (2003). Post-editing. En: Somers, H. (Ed.). *Computers and translation: A Translator's Guide*. (pp. 297-318). Amsterdam/Philadelphia: John Benjamins.
- Alonso, E., y Nunes-Vieira, L. (2017). The Translator's Amanuensis 2020. *The Journal of Specialised Translation*, 28, 345-361.
- ALPAC. (1966). *Language and Machines — Computers in Translation and Linguistics* (A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences). Washington, DC: National Academy of Sciences, National Research Council.
- Al-Shedivat, M., y Parikh, A. (2019). Consistency by Agreement in Zero-shot Neural Machine Translation. *Proceedings of NAACL*. Recuperado de: <https://arxiv.org/abs/1904.02338>
- Amaral, F., Tiago, T. y Tiago, F. (2014). User-Generated Content: Tourists' Profiles on TripAdvisor. *International Journal on Strategic Innovative Marketing* . 1, 3, 137-147. Recuperado de: [https://pdfs.semanticscholar.org/983d/babf4a40093e9d24629eb5f6708e3561f5eb.pdf?\\_ga=2.63621097.772943026.1572872870-578501409.1568740183](https://pdfs.semanticscholar.org/983d/babf4a40093e9d24629eb5f6708e3561f5eb.pdf?_ga=2.63621097.772943026.1572872870-578501409.1568740183)
- Anastasiou, D., y Schäler, R. (2010). Translating Vital Information: Localisation, Internationalisation, and Globalisation. *Syn-Thèses*, 3, 11-25. Recuperado de: <https://ejournals.lib.auth.gr/syn-theses/article/view/5150/5049>
- Androutsopoulos, J. (2006a). Introduction: Sociolinguistics and Computer-Mediated Communication. *Journal of Sociolinguistics*, 10(4): 419–38.
- Aranberri, N. (2014). Postedició, productivitat i qualitat. *Revista Tradumàtica: tecnologies de la traducció*, (12), 471-477.
- Arealillo-Doval, J. J. (2000). La Localización: concepto, nuevas tecnologías y requisitos. En C. Valero Garcés, y I. de la Cruz Cabanillas, Traducción y nuevas tecnologías, herramientas auxiliares del traductor: encuentros en torno a la traducción (págs. 115-124). Alcalá de Henares: Universidad de Alcalá de Henares, Servicio de Publicaciones.
- Arivazhagan, N. et al. (2019). The Missing Ingredient in Zero-Shot Neural Machine Translation. *arXiv preprint arXiv:1903.07091*. Recuperado de: <https://arxiv.org/abs/1903.07091>

- Armentano-Oller, C. et al. (2007). Apertium, una plataforma de código abierto para el desarrollo de sistemas de traducción automática. En: *Proceedings of the FLOSS International Conference 2007*, 5–20. Servicio de Publicaciones de la Universidad de Cádiz.
- Ataman, D., Di Gangi, M. A., y Federico, M. (2018). Compositional Source Word Representations for Neural Machine Translation. En: *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Universitat d'Alacant, Alacant, España, pp. 31-40.
- Bahdanau, D., Cho, K., y Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv preprint arXiv:1409.0473*. Recuperado de: <https://arxiv.org/abs/1409.0473>
- Banerjee, S. y Chua, A. (2014). A Linguistic framework to Distinguish Between Genuine and Deceptive Online Reviews. *Proceedings of the IAENG International Conference on Internet Computing and Web Services* (pp. 501-506). Hong Kong. Recuperado de: [http://www.iaeng.org/publication/IMECS2014/IMECS2014\\_pp501-506.pdf](http://www.iaeng.org/publication/IMECS2014/IMECS2014_pp501-506.pdf)
- Bar-Hillel, Y. (1951). The Present State of Research on Mechanical Translation. *American Documentation*. Recuperado de: <https://doi.org/10.1002/asi.5090020408>
- Bar-Hillel, Y. (1960). The Present Status of Automatic Translation of Languages. *Advances in Computers*, 1(C), 91–163. Recuperado de: [https://doi.org/10.1016/S0065-2458\(08\)60607-5](https://doi.org/10.1016/S0065-2458(08)60607-5)
- Barreda, A. y Bilgihan, A. (2013). An Analysis of User-Generated Content for Hotel Experiences. *Journal of Hospitality and Tourism Technology*. (4)3, 263-280. Recuperado de: <https://doi.org/10.1108/JHTT-01-2013-0001>
- Bayatli, S. et al. (2018). Rule-Based Machine Translation from Kazakh to Turkish. En: *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Universitat d'Alacant, Alacant, España, pp. 49-58.
- Bentivogli, L. et al. (2016). Neural Versus Phrase-Based Machine Translation Quality: A Case Study. *arXiv preprint arXiv:1608.04631*. Recuperado de: <https://arxiv.org/abs/1608.04631>
- Berger, A. L. et al. (1994). The Candide System for Machine Translation. In *Proceedings of the Workshop on Human Language Technology* (pp. 157-162). Association for Computational Linguistics.
- Bermúdez-Bausela, M. (2004). El papel del traductor en el mundo profesional actual: el caso del localizador de software. IV Jornadas sobre la formación y profesión del traductor e intérprete (pp. 25 -27). Villaviciosa de Odón: Universidad Alfonso X el Sabio. Recuperado de:

- [http://abacus.universidadeuropea.es/bitstream/handle/11268/6368/2\\_bermudez\\_art.pdf?sequence=2&isAllowed=y](http://abacus.universidadeuropea.es/bitstream/handle/11268/6368/2_bermudez_art.pdf?sequence=2&isAllowed=y)
- Bernal-Merino, M. Á. (2011). A Brief History of Game Localisation. *TRANS (Revista de Traductología)*, 15, p11-p17. Recuperado de: [http://www.trans.uma.es/pdf/Trans\\_15/11-17.pdf](http://www.trans.uma.es/pdf/Trans_15/11-17.pdf)
- Bernal-Merino, M. Á. (2013). *The Localisation of Video Games*. (Tesis doctoral). London: Imperial College. Recuperado de: <https://pdfs.semanticscholar.org/01b5/3b858c0ae8e224862b5c2d05613a3796f61d.pdf>
- Blackshaw, P. y Nazzaro, M. (2006). *Consumer-generated Media (CGM). Word-of-Mouth in The Age of the Web-fortified Consumer*. New York: Nielsen Buzz Metrics.
- Boyd, D. y Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1): article 11.
- Boyd, D. y Heer, J. (2006). Profiles As Conversation: Networked Identity Performance on Friendster. En: *Proceedings of the Hawai'i International Conference on System Sciences (HICSS-39)*, Kauai, HI, IEEE Computer Society.
- Britannica, T. Editores de Enciclopedia (13 octubre 2023). *Enigma. Encyclopedia Britannica*. <https://www.britannica.com/topic/Enigma-German-code-device>
- Britz, D. et al. (2017). Massive Exploration of Neural Machine Translation Architectures. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1442–1451. Recuperado de: <https://doi.org/10.18653/v1/D17-1151>
- Brown, P. et al. (1988). A Statistical Approach to Language Translation. En: *Proceedings of the 12th Conference on Computational Linguistics. Volume 1* (pp. 71-76). Association for Computational Linguistics.
- Buhalis, D. y Law, R. (2008). Progress in Information Technology and Tourism Management: 20 Years on and 10 Years After the Internet – The State of eTourism Research. *Tourism Management* (29)4, pp. 609-623.
- Cadieux, P., y Esselink, B. (2004). GILT: Globalization, Internationalization, Localization, Translation. *Globalization insider*, 11(1.5), 1-5.
- Cámara de la Fuente, L. (2005). *Comunicación técnica multilingüe: gestión del conocimiento y de recursos lingüísticos para proyectos de traducción*. (Tesis doctoral). Barcelona: Universitat Pompeu Fabra.
- Cancedda, N. et al. (2009). A Statistical Machine Translation Primer. En: *Learning Machine Translation* (pp. 1–37). Recuperado de: <https://ieeexplore.ieee.org/document/6279396>

- Candel-Mora, M. Á., y Borja-Tormo, C. (2017). Desarrollo de la aplicación Post-editing Calculeffort para la estimación del esfuerzo en posesición. *Tradumàtica*, (15), 1-9.
- Candel-Mora, M. Á. (2015). Evaluation of English to Spanish MT Output of Tourism 2.0 Consumer-Generated Reviews with Post-Editing Purposes. En: *Translating and the Computer* 37. (37 - 47). London, UK: Editions Tradulex, Geneva
- Carl, M. y Way, A. (eds). (2003). *Recent Advances in Example-Based Machine Translation*. Text, Speech and Language Technology. Springer Netherlands.
- Casacuberta, F., y Peris-Abril, Á. (2017). Traducció automàtica neuronal. *Revista Tradumàtica* (15), 66-74. Recuperado de: <https://revistes.uab.cat/ojs-tradumatica/tradumatica/article/view/n15-casacuberta-peris>
- Castilho, S. et al. (2017). Is Neural Machine Translation the New State of the Art?. *The Prague Bulletin of Mathematical Linguistics*, 108(1), 109-120. Recuperado de: <https://content.sciendo.com/view/journals/pralin/108/1/article-p109.xml>
- Caswell, I. (2022, Mayo). Google Translate Learns 24 New Languages. Google Blog – The Keyword. <https://blog.google/products/translate/24-new-languages/>
- Cebrián, M. (2008). La Web 2.0 como red social de comunicación e información. *Estudios sobre el mensaje periodístico*, 14, 345-361.
- Chatzisavva, P. (2013). Digital Transformation in Tourism Sector. (Trabajo de fin de máster). Thessaloniki: International Hellenic University. Recuperado de: [https://repository.ihu.edu.gr/xmlui/bitstream/handle/11544/29204/Dissertation\\_Chatzisavva\\_Panagiota.pdf?sequence=1](https://repository.ihu.edu.gr/xmlui/bitstream/handle/11544/29204/Dissertation_Chatzisavva_Panagiota.pdf?sequence=1)
- Chevalier, J. y Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. *Quantitative Marketing and Economics*. 1, 2, 203-222. Recuperado de: <https://faculty.chicagobooth.edu/austan.goolsbee/research/amzn.pdf>
- Chik, A., y Vásquez, C. (2016). A comparative Multimodal Analysis of Restaurant Reviews from Two Geographical Contexts. *Visual Communication*, 16(1), 3-26.
- Cho, K., Van Merriënboer, B., Bahdanau, D., y Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv preprint arXiv:1409.1259*.
- Clemons, E., Gao, G., y Hitt, L. (2006). When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry. *Journal of Management Information Systems*, 23(2), 149-171. Recuperado de: <https://www.tandfonline.com/doi/abs/10.2753/MIS0742-1222230207>
- Comisión Europea. (2014). Study on Online Consumer Reviews in the Hotel Sector. Recuperado de: <http://bookshop.europa.eu/en/study-on-onlineconsumer-reviews-in-the-hotel-sector-pbND0414464/>



- Cooper, C. y Hall, C. M. (2007). *Contemporary Tourism: An International Approach*. Gran Bretaña, Routledge.
- Corte, N. (2002). Localización e internacionalización de sitios web. *Revista Tradumática*, 1, 1-8. Recuperado de: <http://www.fti.uab.es/tradumatica/revista/articles/ncorte/ncorte.PDF>
- Costa-Jussa, M. R. et al. (2012). Study and Comparison of Rule-Based and Statistical Catalan-Spanish Machine Translation Systems. *Computing and Informatics*, 31(2), 245-270.
- Craciunescu, O., Gerding-Salas, C., & Stringer-O'Keeffe, S. (2004). Machine Translation and Computer-Assisted Translation. *Machine Translation and Computer-Assisted Translation*.
- Dandapat, S. y Lewis, W. (2018). Training Deployable General Domain MT for a Low Resource Language Pair: English–Bangla. En: *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Universitat d'Alacant, Alacant, España, pp. 109-117.
- Darics, E. (2015). *Digital Business Discourse*. Londres, Gran Bretaña, PALGRAVE MACMILLAN.
- Dellarocas, C. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, 49(10), 1407-1427.
- Dellarocas, C. et al. (2004) "Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning." *Proceedings of the 24th International Conference on Information Systems*, Washington D.C.
- Destination Analysts. (2019). The Future of Travel-Planning Apps. En *The State of the American Traveler, Travel Outlook*. Recuperado de: <https://www.destinationanalysts.com/welcome-to-2019-american-traveler-sentiment-weakens-2-2/>
- Dhar, V. y Chang, E. (2009). Does Chatter Matter? The Impact of User-Generated Content on Music Sales. *Journal of Interactive Marketing*. 23, 4, 300-307.
- Díaz-Montón, D. (2006). Entresijos de la localización. Meristation, pp. 1-4. Recuperado de: <http://isg.urv.es/library/papers/isgbook.pdf>
- Diéguez-Morales, M. I. (2008). Análisis terminológico de sitios web localizados del inglés al español: uso de técnicas de amplificación y elisión (1). *Revista Tradumática*, 6, 1-22. Recuperado de: <http://www.fti.uab.es/tradumatica/revista/num6/articles/09/9.pdf>
- Do Carmo, F. (2018, May). Does Machine Translation Really Produce Translations?. In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*.
- Doherty, S. (2017). Issues in Human and Automatic Translation Quality Assessment. *Human Issues in Translation Technology*, 131-148.

- Dostert, L. E. (1957). Brief History of Machine Translation Research. En *Research in Machine Translation*.
- Dugast, L., Senellart, J., y Koehn, P. (2007). Statistical Post-Editing on SYSTRAN's Rule-Based Translation System. En: *Proceedings of the Second Workshop on Statistical Machine Translation* (pp. 220-223).
- Egbert, J. y Biber, D. (2013). Developing a User-Based Method of Web Register Classification. En: S. Evert, E. Stemle y P. Rayson (eds). *Proceedings of the 8th Web as Corpus Workshop (WAC-8) @Corpus Linguistics 2013*, pp. 16–23.
- El contenido generado por usuarios, clave en la estrategia de marketing. (5 de abril 2018). Hosteltur. Recuperado de: [https://www.hosteltur.com/127459\\_contenido-generado-usuarios-clave-estrategia-marketing.html](https://www.hosteltur.com/127459_contenido-generado-usuarios-clave-estrategia-marketing.html)
- España-Bonet, C., y Costa-Jussa, M. R. (2016) Hybrid Machine Translation Overview. En: Costa-Jussa M., Rapp R., Lambert P., Eberle K., Banchs R., Babych B. (eds) *Hybrid Approaches to Machine Translation. Theory and Applications of Natural Language Processing* (pp. 1-24). Springer, Cham
- Esselink, B. (2000). A Practical Guide to Localization (Vol. 4). Amsterdam-Philadelphia: John Benjamins Publishing.
- Esselink, B. (2002). Localization Engineering: The Dream Job? *Revista Tradumàtica*, 1, 1-5.  
Recuperado de:  
<http://www.fti.uab.es/tradumatica/revista/articles/besselink/besselink.PDF>
- Esselink, B. (2003). The Evolution of Localization. *Multilingual Computing y Technology*, 14 (5), 4-7.
- Estellés, A. (2010). Revisión del entorno de trabajo de un traductor de contenido web. *Revista Tradumàtica*, 8, 1-12. Recuperado de:  
[http://revistes.uab.cat/tradumatica/article/viewFile/n8-estelles/pdf\\_18](http://revistes.uab.cat/tradumatica/article/viewFile/n8-estelles/pdf_18)
- Fernández-Castrillo, C. (2014). Prácticas transmedia en la era del prosumidor: Hacia una definición del Contenido Generado por el Usuario (CGU). *CIC. Cuadernos de información y comunicación*, 19, 53-67. Recuperado de:  
<https://revistas.ucm.es/index.php/CIYC/article/view/43903/41501>
- Fernández-Costales, A. (2009). Translation 2.0. The Localization of Institutional Websites Under the Scope of Functionalist Approaches. *CETRA Research Papers 2008*, 1-18.
- Fernández-Costales, A. (2010). *Traducción, Localización e Internacionalización. El caso de las Páginas Web Universitarias*. (Tesis doctoral). Oviedo: Universidad de Oviedo.  
Recuperado de:  
<https://www.educacion.gob.es/teseo/imprimirFicheroTesis.do?idFichero=19611>

- Foley, B. (2023). Tripadvisor Review Transparency Report. Tripadvisor. Recuperado de: <https://www.tripadvisor.com/TransparencyReport2023>
- Forcada, M. L. (2009). Apertium: traducció automàtica de codi obert per a les llengües romàniques. *Linguamàtica*, 1(1), 13-23.
- Forcada, M. L. (2017). Making Sense of Neural Machine Translation. *Translation spaces*, 6(2), 291-309. Recuperado de: <https://www.jbe-platform.com/content/journals/10.1075/ts.6.2.06for>
- Forcada, M. L., Tyers, F., y Ramírez-Sánchez, G. (2009). The Apertium Machine Translation Platform: Five Years on. In *Proceedings of the First International Workshop on Free/Open-Source Rule-Based Machine Translation* (pp. 3-10).
- Fry, D. (2003). The Localization Industry Primer. (Revised by Arle Lommel). *Localization and Standards Association*.
- Gao, G., Gu, B., y Lin, M. (2007). The Dynamics of Online Consumer Reviews. En: *Workshop on Information Systems and Economics (WISE)*.
- García de Torres, E. (2010). Contenido generado por el usuario: aproximación al estado de la cuestión. *El profesional de la información*. 19, 6, 585-594. Recuperado de: <http://www.elprofesionaldeinformacion.com/contenidos/2010/noviembre/04.pdf>
- García-Nevarés, L. A. (2016). *Localización e internacionalización de software: puntos de encuentro entre el localizador y el programador*. (Tesis doctoral). Salamanca: Universidad de Salamanca. Recuperado de: [https://gedos.usal.es/jspui/bitstream/10366/132948/1/DTI\\_Garc%C3%ADaNevarésLA\\_Localizaci%C3%B3nInternalizaci%C3%B3n.pdf](https://gedos.usal.es/jspui/bitstream/10366/132948/1/DTI_Garc%C3%ADaNevarésLA_Localizaci%C3%B3nInternalizaci%C3%B3n.pdf)
- García-Saavedra Valle, M. T. (s.f.). ¿Traduces o localizas? La localización, futuro y presente de la traducción. Acta Congreso de Almagro, Congresos Internacionales de Esletra, (pp. 409-423). Almagro. Recuperado de: [https://cvc.cervantes.es/lengua/esletra/pdf/01/029\\_garcia.pdf](https://cvc.cervantes.es/lengua/esletra/pdf/01/029_garcia.pdf)
- Gartner, W. C. (1993). Image Formation Process. *Journal of Travel and Tourism Marketing*. 2, 2/3, 191-215. Recuperado de: [https://www.researchgate.net/publication/254379044\\_Image\\_Formation\\_Process](https://www.researchgate.net/publication/254379044_Image_Formation_Process)
- Gaspari, F., y Hutchins, W. J. (2007). Online And Free! Ten Years of Online Machine Translation: Origins, Developments, Current Use and Future Prospects. In *Proceedings of Machine Translation Summit XI: Papers*, 199-206.
- Gavaldà, M. (2000). La investigación en tecnologías de la lengua. *Quark*, (19), 21-25.
- Gehring, J. et al. (2016). A Convolutional Encoder Model for Neural Machine Translation. *arXiv preprint arXiv:1611.02344*. Recuperado de: <https://arxiv.org/abs/1611.02344>

- Gomes de Oliveira, R., y Anastasiou, D. (2011). Comparison of SYSTRAN and Google Translate for English→ Portuguese. *Revista Tradumàtica*, (9), 118-136.
- Gordin, M. D. (2015). *Scientific Babel: How Science Was Done Before and After Global English*. University of Chicago Press.
- Görög, A. y Sánchez-Jijón, P. (2014). Translation and Quality. *Revista Tradumàtica*, 12, 388-391.  
Recuperado de:  
[https://ddd.uab.cat/pub/tradumatica/tradumatica\\_a2014n12/tradumatica\\_a2014n12\\_p388iSPA.pdf](https://ddd.uab.cat/pub/tradumatica/tradumatica_a2014n12/tradumatica_a2014n12_p388iSPA.pdf)
- Gotti, F., Langlais, P., y Lapalme, G. (2014). Designing a Machine Translation System for Canadian Weather Warnings: A Case Study. *Natural Language Engineering*, 20(3), 399-433.
- Gough, N., y Way, A. (2004). Robust Large-Scale EBMT with Marker-Based Segmentation. En: *TMI 2004 - 10th International Conference on Theoretical and Methodological Issues in Machine Translation*. Baltimore, Maryland, USA.
- Gretzel, U. et al. (2007). Online Travel Review Study: Role & Impact of Online Travel Reviews. Laboratory for Intelligent Systems in Tourism, Texas A&M University. Recuperado de:  
<https://www.tripadvisor.com/pdfs/OnlineTravelReviewReport.pdf>
- Guizzo, E. M. (2003). *The Essential Message: Claude Shannon and the Making of Information Theory*. (Tesis doctoral). *Diss. Massachusetts Institute of Technology*.
- Ha, T. L., Niehues, J., y Waibel, A. (2017). Effective Strategies in Zero-Shot Neural Machine Translation. *arXiv preprint arXiv:1711.07893*. Recuperado de:  
<https://arxiv.org/abs/1711.07893>
- Henisz-Dostert, B., Macdonald, R. R., y Zarechnak, M. (2011). *Machine Translation* (Vol. 11). Walter de Gruyter.
- Hennig-Thurau, T., et al. (2004) Electronic Word-Of-Mouth Via Consumer Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet? *Journal of Interactive Marketing* 18: 38–52.
- Hermida, A. y Thurman, N. (2008). A Clash of Cultures. The Integration of User-Generated Content Within Professional Journalistic Frameworks at British Newspaper Websites. *Journalism practice*. (2)3, 343-356.
- Herring, S. C. (2012). Grammar and Electronic Communication. En: C. A. Chapelle (Ed.), *Encyclopedia of Applied Linguistics*. Hoboken, NJ: Wiley/Blackwell.
- Herring, S. C. et al. (2005). Conversations in the Blogosphere: An Analysis “From the Bottom Up”. En: *Proceedings of the 38th Hawaii International Conference on System Sciences*, Kauai, HI, IEEE Computer Society.

- Hidalgo-Alcázar, C., Sicilia, M. y Ruíz, S. (2015). La Imagen de un Producto Turístico Rural a través del Acceso al Contenido Generado por otros Usuarios en Internet: Diferencias por Género. *Journal of Technology Management & Innovation*. 10, 3, 75-84. Recuperado de: <https://scielo.conicyt.cl/pdf/jotmi/v10n3/art09.pdf>
- Hutchins, J. (1986). *Machine Translation: Past, Present, Future*. Chichester: England, Ellis Horwood.
- Hutchins, J. (1995). Machine Translation: A Brief History. En *Concise history of the language sciences* (pp. 431-445). Pergamon.
- Hutchins, J. (2001). Machine Translation and Human Translation: In Competition or in Complementation. *International Journal of Translation*, 13 (1-2), 5-20.
- Hutchins, J. (2001). Machine Translation Over Fifty Years. *Histoire épistémologie langage*, 23(1), 7-31.
- Hutchins, J. (2003). ALPAC: The (in) Famous Report. *Readings in Machine Translation*, 14, 131-135.
- Hutchins, J. (2003). The Development and Use of Machine Translation Systems and Computer-based Translation Tools. En: *International Symposium on Machine Translation and Computer Language Information Processing*, Beijing, China.
- Hutchins, J. (2004). Two Precursors of Machine Translation: Artsrouni and Trojanskij. *International Journal of Translation*, 16(1), 11-31.
- Hutchins, J. (2005). Example-Based Machine Translation: A Review and Commentary. *Machine Translation*, 19(3-4), 197-211.
- Hutchins, J. (2005). The History of Machine Translation in a Nutshell. Recuperado de: <http://www.hutchinsweb.me.uk/Nutshell-2014.pdf>
- Hutchins, J. (2005). Towards a Definition of Example-Based Machine Translation. En *Machine Translation Summit X, Second Workshop on Example-Based Machine Translation* (pp. 63-70) Phuket, Thailand.
- Hutchins, J. (2007). Machine Translation: A Concise History. *Computer Aided Translation: Theory and practice*, 13(29-70), 11. Recuperado de: <https://pdfs.semanticscholar.org/e97a/40cc28ce17a17ce9b73d77e69ffa1210fa25.pdf>
- Hutchins, J. (2015). Machine Translation: History of Research and Applications. *Routledge Encyclopedia of Translation Technology*. Recuperado de: <http://www.hutchinsweb.me.uk/Routledge-2014.pdf>
- Jean, S., Cho, K., Memisevic, R., y Bengio, Y. (2014). On Using Very Large Target Vocabulary for Neural Machine Translation. *arXiv preprint arXiv:1412.2007*. Recuperado de: <https://arxiv.org/abs/1412.2007>

- Jiménez-Crespo, M. A. (2008). *El proceso de localización web: estudio contrastivo de un corpus comparable del género sitio web corporativo*. (Tesis doctoral). Granada: Universidad de Granada. Recuperado de: <https://hera.ugr.es/tesisugr/17515324.pdf>
- Jiménez-Crespo, M. A. (2009). Conventions in Localisation: A Corpus Study of Original Vs. Translated Web Texts. *Jostrans: The Journal of Specialized Translation*, 12, 79-102. Recuperado de: [http://www.jostrans.org/issue12/art\\_jimenez.php](http://www.jostrans.org/issue12/art_jimenez.php)
- Jiménez-Crespo, M. A. (2009). El uso de corpus textuales en localización. *Revista Tradumàtica*, 7, 1-15. Recuperado de: <http://www.fti.uab.es/tradumatica/revista/num7/articles/05/05.pdf>
- Johnson, M. et al. (2017). Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics*, 5, 339-351. Recuperado de: [https://www.mitpressjournals.org/doi/abs/10.1162/tacl\\_a\\_00065](https://www.mitpressjournals.org/doi/abs/10.1162/tacl_a_00065)
- Johnson, M., Firat, O., y Aharoni, R. (2019). Massively Multilingual Neural Machine Translation. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 3874–3884. Recuperado de: <https://www.aclweb.org/anthology/N19-1388>
- Jones, R. H., Chik, A., y Hafner, C. A. (2015). *Discourse and Digital Practices*. New York y Londres, Routledge.
- Kalchbrenner, N. et al. (2016). Neural Machine Translation in Linear Time. *arXiv preprint arXiv:1610.10099*. Recuperado de: <https://arxiv.org/abs/1610.10099>
- Kay, M. (1997). The Proper Place of Men and Machines in Language Translation. *Machine Translation*, 12(1-2), 3-23.
- Kelly, N. (3 de agosto de 2012). Speak to Global Customers in their Own Language. *Harvard Business Review*. Recuperado de: <https://hbr.org/2012/08/speak-to-global-customers-in-t/>
- Koletnik, M. (2011). Applicability and Challenges of Using Machine Translation in Translator Training. *ELOPE: English Language Overseas Perspectives and Enquiries*, 8(2), 7-18.
- Kreck, L. (1998). Service Quality: Who Should Determine It? Research and Practice. *Journal of International Hospitality, Leisure & Tourism Management*, (1)4, 63-77.
- Krumm, J., Davies, N. y Narayanaswami, C. (2008). User Generated Content. *IEEE Pervasive Computing*. 7, 4, 10-11. Recuperado de: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4653465>
- Lakew, S. M. et al. (2018). Improving Zero-Shot Translation of Low-Resource Languages. *arXiv preprint arXiv:1811.01389*. Recuperado de: <https://arxiv.org/abs/1811.01389>

- Lamb, A. y Xie, M. (2016). Convolutional Encoders for Neural Machine Translation. Recuperado de: <https://cs224d.stanford.edu/reports/LambAndrew.pdf>
- Langlais, P., et al. (2005). The Long-Term Forecast for Weather Bulletin Translation. *Machine Translation*, 19, 83-112.
- Lappas, T. et al. (2016). The Impact of Fake Reviews on Online Visibility: A Vulnerability Assessment of the Hotel Industry. *Information Systems Research*, 27(4), 940-961. Recuperado de: <https://pubsonline.informs.org/doi/abs/10.1287/isre.2016.0674>
- Li, X., y Hitt, L. (2004). Self Selection and Information Role of Online Product Reviews. *Workshop on Information Systems and Economics*. Washington, DC.
- Libovický, J., Brovelli, M.T. y Cartoni, B. (2018). Machine Translation Evaluation Beyond the Sentence Level. En: Pérez-Ortiz, J. A., et al. (Eds.). *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Universitat d'Alacant, Alacant, España, pp. 179-188. Recuperado de: <http://rua.ua.es/dspace/handle/10045/76047>
- LISA (1998). LEIT Phase I Final Report. LISA Education Initiative Taskforce (LEIT), Report November, 22.
- Lison, P. (2016). *Introduction to Statistical Machine Translation* [Presentación en Power Point]. *Maskinlæring meetup*, Oslo, Noruega. Recuperado de: [http://folk.uio.no/plison/pdfs/talks/meetup\\_SMT.pdf](http://folk.uio.no/plison/pdfs/talks/meetup_SMT.pdf)
- Litvin, S. W., Goldsmith, R. E. y Pan, B. (2008). Electronic Word-Of-Mouth in Hospitality and Tourism Management. *Tourism Management*. 29, 3, 458-468. Recuperado de: [https://www.researchgate.net/publication/222394083\\_Electronic\\_Word-of-Mouth\\_in\\_Hospitality\\_and\\_Tourism\\_Management](https://www.researchgate.net/publication/222394083_Electronic_Word-of-Mouth_in_Hospitality_and_Tourism_Management)
- Locke, W. N., y Booth, A. D. (1955). Historical Introduction. En *Machine Translation of Languages: Fourteen Essays*, ed. A. D. Booth and W. N. Locke, 1–14. Cambridge, MA: MIT Press.
- Lommel, A. R. y DePalma, D. A. (2016). Europe's Leading Role in Machine Translation. *Common Sense Advisory*, Boston, USA.
- Lommel, A., Uszkoreit, H., y Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Tradumàtica*, (12), 455-463.
- López, A. (2008). Statistical Machine Translation. *ACM Computing Surveys (CSUR)*, 40(3), 1-49.
- Luong, M. T., Pham, H., y Manning, C. D. (2015). Effective Approaches to Attention-Based Neural Machine Translation. *arXiv preprint arXiv:1508.04025*. Recuperado de: <https://arxiv.org/abs/1508.04025>

- Mackiewicz, J. (2008). Reviewer Motivations, Bias, and Credibility in Online Reviews. En: S. Kelsey y K. St. Amant (eds.) *Handbook of Research on Computer Mediated Communication* (pp. 252–66). Hershey, PA, The Idea Group Publishers.
- Manap, K. A., y Adzharudin, N. A. (2013). The Role of User Generated Content (UGC) in Social Media for Tourism Sector. En: *The 2013 WEI international academic conference proceedings* (pp. 52-58).
- Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*. 43, 3, 345-354
- McElhaney, T., y Vasconcellos, M. (1988). The Translator and the Postediting Experience. *Technology as translation strategy*, 2, 140-148.
- Mudambi, S. M., y Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly: Management Information Systems*, 34(1), 185–200. Recuperado de: <https://doi.org/10.2307/20721420>
- Muller, P. et al. (2011). Consumer Behaviour an a Digital Environment. *Policy Department Economic and Scientific Policy, European Parliament, B-1047 Brussels*. Recuperado de: <https://www.europarl.europa.eu/document/activities/cont/201108/20110825ATT25258/20110825ATT25258EN.pdf>
- Munar, A. M. et al. (2013). *Tourism Social Media Transformations in Identity Community and Culture*. Bradford, Gran Bretaña, Emerald Group Publishing Limited.
- Nafria, I. (2007). *Web 2.0: El usuario, el nuevo rey de Internet*. Barcelona: Gestión 2000.
- Ney, H. (2004). Statistical Machine Translation and its Challenges. En: *Eighth International Conference on Spoken Language Processing*.
- O'Brien, S., Choudhury, R., van der Meer, J. y Aranberri, N. (2011). "Translation Quality Evaluation Framework". TAUS Report.
- Och, F. J. (2005). Statistical Machine Translation: Foundations and Recent Advances. *Tutorial at MT Summit*. Phuket, Thailand.
- O'Hagan, M. y Ashworth, D. (2002). Translation-Mediated Communication in a Digital World. Facing the Challenges of Globalization and Localization. Clevedon: Multilingual Matters LTD.
- Oliver, A. (2014). *Traducció i tecnologies: eines, processos i recursos*. Barcelona, Universitat Oberta de Catalunya.
- Ong, E., Go, K., Morga, M., Nunez, V., y Veto, F. (2007). Extracting and Using Translation Templates in an Example-Based Machine Translation System. *Journal of Research in Science, Computing, and Engineering*, 4(3), 83-100.



- Ong, T., Mannino, M., y Gregg, D. (2014). Linguistic Characteristics of Skill Reviews. *Electronic Commerce Research and Applications*, 13(2), 69-78. Recuperado de: <https://www.sciencedirect.com/science/article/pii/S1567422313000665>
- Ortega, J. E., Lu, W., Meyers, A., y Cho, K. (2018). Letting a Neural Network Decide Which Machine Translation System to Use for Black-Box Fuzzy-Match Repair. En: *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Universitat d'Alacant, Alacant, España, pp. 209-218.
- Pagans, M. (2002). Localització, ens ubiquem? *Revista Tradumàtica*, 1, 1-4. Recuperado de: <https://ddd.uab.cat/pub/tradumatica/15787559n1/15787559n1a1pagans.pdf>
- Pan, B., Maclaurin, T. y Crofts, J. (2007). Travel Blogs and Their Implications for Destination Marketing. *Journal of Travel Research*. 46, 1, 35-45. Recuperado de: [https://www.researchgate.net/publication/228838372\\_Travel\\_Blogs\\_and\\_Their\\_Implications\\_for\\_Destination\\_Marketing](https://www.researchgate.net/publication/228838372_Travel_Blogs_and_Their_Implications_for_Destination_Marketing)
- Pérez-Cabrera, M. J. (2012). Aspectos técnicos y metodológicos de la localización de contenido web. *Odisea*, 13, 189-207. Recuperado de: [http://repositorio.ual.es/bitstream/handle/10835/2089/Odisea13\\_Perez.pdf?sequence=1](http://repositorio.ual.es/bitstream/handle/10835/2089/Odisea13_Perez.pdf?sequence=1)
- Pérez-Fernández, L. M. (2010). *La localización de videojuegos (inglés-español): aspectos técnicos, metodológicos y profesionales*. (Tesis doctoral). Málaga: Universidad de Málaga.
- Peris-Abril, Á., Domingo-Ballester, M., y Casacuberta, F. (2017). Interactive Neural Machine Translation. *Computer Speech & Language*, 45, 201-220.
- Peris-Abril, Á., y Casacuberta, F. (2015). A Bidirectional Recurrent Neural Language Model for Machine Translation. *Procesamiento del Lenguaje Natural*, (55), 109-116. Recuperado de:
- Pike, S. y Ryan, C. (2004). Destination Positioning Analysis Through a Comparison of Cognitive, Affective and Cognitive Perceptions. *Journal of Travel Research*. 42, 2, 333-342. Recuperado de: [https://www.researchgate.net/publication/27475984\\_Destination\\_Positioning\\_Analysis\\_through\\_a\\_Comparison\\_of\\_Cognitive\\_Affective\\_and\\_Cognitive\\_Perceptions](https://www.researchgate.net/publication/27475984_Destination_Positioning_Analysis_through_a_Comparison_of_Cognitive_Affective_and_Cognitive_Perceptions)
- Plaza-Lara, C. (2014). *La enseñanza de la traducción asistida por ordenador: análisis didáctico-curricular y propuesta de integración de la subcompetencia instrumental profesional en el aula*. (Tesis doctoral). Málaga: Universidad de Málaga. Recuperado de: <https://riuma.uma.es/xmlui/handle/10630/8126>

- Pollach, I. (2006). Electronic Word of Mouth: A Genre Analysis of Product Reviews on Consumer Opinion Web Sites. En: *Proceedings of the 39th Hawaii International Conference on System Sciences*. IEEE Computer Society. Recuperado de: [https://www.researchgate.net/publication/4216119\\_Electronic\\_Word\\_of\\_Mouth\\_A\\_Genre\\_Analysis\\_of\\_Product\\_Reviews\\_on\\_Consumer\\_Opinion\\_Web\\_Sites](https://www.researchgate.net/publication/4216119_Electronic_Word_of_Mouth_A_Genre_Analysis_of_Product_Reviews_on_Consumer_Opinion_Web_Sites)
- Poncelas, A. et al. (2018). Investigating Backtranslation in Neural Machine Translation. En: *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Universitat d'Alacant, Alacant, España, pp. 249-258.
- Pouget-Abadie, J., Bahdanau, D., Van Merriënboer, B., Cho, K., y Bengio, Y. (2014). Overcoming The Curse of Sentence Length for Neural Machine Translation Using Automatic Segmentation. *arXiv preprint arXiv:1409.1257*.
- Presi, C., Saridakis, C. y Hartmans, S. (2014). User-Generated Content Behaviour of the Dissatisfied Service Customer. *European Journal of Marketing*. 48, 9/10, 1600-1625. Recuperado de: <http://eprints.whiterose.ac.uk/91086/2/EJM-07-2012-0400-R1-Paper-FINAL.pdf>
- Pullman, M. et al. (2005). Let Me Count the Words: Quantifying Open Ended Interactions With Guests. *Cornell Hotel & Restaurant Administration Quarterly*, (46)3, 323-343.
- Puri, A. (2007), "The Web of Insights: The Art and Practice of Webnography", *International Journal of Market Research*, Vol. 49 No. 3, pp. 387-408.
- Purnawirawan, N. et al. (2015). A Meta-Analytic Investigation of the Role of Valence in Online Reviews. *Journal of Interactive Marketing*. (31), 17-27. Recuperado de: <https://doi.org/10.1016/j.intmar.2015.05.001>.
- Pym, A. (2010). Web Localization. En K. Malmkjær, y K. Windle, *The Oxford Handbook of Translation Studies*. (págs. 410-424). Oxford: Oxford University Press. Recuperado de: [http://usuaris.tinet.cat/apym/on-line/translation/2009\\_website\\_localization\\_feb.pdf](http://usuaris.tinet.cat/apym/on-line/translation/2009_website_localization_feb.pdf)
- Pym, A. (2012). *Teorías contemporáneas de la traducción (exploraciones pedagógicas)*. Tarragona: Intercultural Studies Group. URV.
- Quah, C. K. (2006). *Translation and Technology*. Basingstoke: Palgrave Macmillan.
- Quirk, C., y Menezes, A. (2006). Dependency Treelet Translation: The Convergence of Statistical and Example-Based Machine-Translation? *Machine Translation*, 20(1), 43-65.
- Ramírez-Polo, L. (2012). *Use and Evaluation of Controlled Languages in Industrial Environments and Feasibility Study for the Implementation of Machine Translation*. (Tesis doctoral). València: Universitat de València. Recuperado de: <https://uv.academia.edu/LauraRamirezPolo>

- Ressin, M. (2015). *An Empirical Examination of Interdisciplinary Collaboration within the Practice of Localisation and Development of International Software*. (Tesis doctoral). London: University of West London. Recuperado de: [http://repository.uwl.ac.uk/id/eprint/2858/1/Malte%20Ressin%20PhD%20thesis%20\(Final%20-%20Oct%2015\).pdf](http://repository.uwl.ac.uk/id/eprint/2858/1/Malte%20Ressin%20PhD%20thesis%20(Final%20-%20Oct%2015).pdf)
- Ricci, F. y Wietsma, R. (2006). Product Reviews in Travel Decision-Making. En: M. Hitz, M. Sigala y J. Murphy (eds.) *Information and Communication Technologies in Tourism* (pp. 296–307). Vienna: Springer.
- Rubio, R., y Bernal, J. (2016). *Introducción a la lingüística computacional*. Bogotá: Ediciones de la U.
- Ruzic, D. y Biloz, A. (2010). *Tourism and Hospitality Management 2010: Conference Proceedings: Social Media in Destination Marketing Organizations (DMO's)*.178-190.
- Sánchez-Gijón, P. (2016). La posesición: hacia una definición competencial del perfil y una descripción multidimensional del fenómeno. *Sendebarr*, 27, 151-162.
- Sánchez-Gijón, P. y Oliver, A. (2018). Traducción automática estadística para traductores. *Taller de traducción automática estadística para traductores*, Madrid. Recuperado de: <https://ddd.uab.cat/record/191254>.
- Sánchez-Gijón, P., Moorkens, J., y Way, A. (2018). Perception vs. Acceptability of TM and SMT Output: What do translators prefer?. En: *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Universitat d'Alacant, Alacant, España, pp. 331.
- Sánchez-Gijón, P., Moorkens, J., y Way, A. (2019). Post-Editing Neural Machine Translation Versus Translation Memory Segments. *Machine Translation*, 33(1-2), 31-59. Recuperado de: <https://link.springer.com/article/10.1007/s10590-019-09232-x>
- Sánchez-Martínez, F. (2008). *Empleo de métodos no supervisados basados en corpus para construir traductores automáticos basados en reglas*. (Tesis doctoral). Alicante: Universidad de Alicante. Recuperado de: [https://rua.ua.es/dspace/bitstream/10045/13879/1/Tesis\\_sanchez.pdf](https://rua.ua.es/dspace/bitstream/10045/13879/1/Tesis_sanchez.pdf)
- Sandrini, P. (2008). Localization and Translation. *MuTra Journal*, 2, 167-191. Recuperado de: <http://homepage.uibk.ac.at/~c61302/publik/localiz.pdf>
- Sato, S., y Nagao, M. (1990). Toward Memory-Based Translation. In *COLING 1990 Volume 3: Papers Presented to the 13th International Conference on Computational Linguistics*.
- Schäler, R. (2008). Linguistic Resources and Localization. In E. Yuste Trigo, *Topics in Language Resources for Translation and Localization*. (pp. 195-225). Ámsterdam-Filadelfia: John Benjamins.

- Schemmann, B. (2011). A Classification of Presentation Forms of Travel and Tourism-Related Online Consumer Reviews. *e-Review of Tourism Research* 2. Recuperado de: <http://ertr.tamu.edu/enter-2011-short-papers>
- Scholand, M. (2002). Localización de Videojuegos. *Revista Tradumàtica*, 1, 1-9. Recuperado de: <http://www.fti.uab.es/tradumatica/revista/articulos/mscholand/mscholand.PDF>
- Schulz, P., Aziz, W., y Cohn, T. (2018). A Stochastic Decoder for Neural Machine Translation. *arXiv preprint arXiv:1805.10844*. Recuperado de: <https://arxiv.org/abs/1805.10844>
- Schuster, M., Johnson, M., y Thorat, N. (2016). Zero-Shot Translation with Google's Multilingual Neural Machine Translation System. *Google AI Blog*, 22. Recuperado de: <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>
- Seneff, S., Wang, C., y Lee, J. (2006). Combining Linguistic and Statistical Methods for Bi-Directional English Chinese Translation in the Flight Domain. En: *Proceedings of AMTA*.
- Sennrich, R., Haddow, B., y Birch, A. (2015). Neural Machine Translation of Rare Words with Subword Units. *arXiv preprint arXiv:1508.07909*. Recuperado de: <https://arxiv.org/abs/1508.07909>
- Seoane-Vicente, A. L. (2015). *Lenguaje controlado aplicado a la traducción automática de prospectos farmacéuticos*. (Tesis doctoral). Alicante: Universidad de Alicante. Recuperado de: [https://rua.ua.es/dspace/bitstream/10045/53587/1/tesis\\_angel\\_luis%20seoane.pdf](https://rua.ua.es/dspace/bitstream/10045/53587/1/tesis_angel_luis%20seoane.pdf)
- Siddhant, A. et al. (2019). Evaluating the Cross-Lingual Effectiveness of Massively Multilingual Neural Machine Translation. *arXiv preprint arXiv:1909.00437*. Recuperado de: <https://arxiv.org/abs/1909.00437>
- Slocum, J. (1985). A Survey of Machine Translation: Its History, Current Status and Future Prospects. *Computational Linguistics*, 11(1), 1-17.
- Smith, A. N., Fischer, E., y Yongjian, C. (2012). How Does Brand-Related User-Generated Content Differ Across Youtube, Facebook, and Twitter? *Journal of Interactive Marketing*, 26(2), 102-113.
- Somers, H. (2003). The Translator's Workstation. *BENJAMINS TRANSLATION LIBRARY*, 35, 13-30.
- Sparks, B. A. y Browning, V. (2010). Complaining in Cyberspace: The Motives and Forms of Hotel Guests' Complaints Online. *Journal of Hospitality Marketing & Management*, (19)7, 797-818. Recuperado de: <https://www.tandfonline.com/doi/abs/10.1080/19368623.2010.508010>
- Sreelekha, S. (2017). Statistical Vs Rule Based Machine Translation; A Case Study on Indian Language Perspective. *arXiv preprint arXiv, 1708*.

- Strandell, H. y Wolff, P. (2019). Key Figures on Europe — Statistics Illustrated. Recuperado de: <https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-EI-19-001>
- Stringam, B. B. y Gerdes, J. Jr (2010). An Analysis of Word-of-Mouse Ratings and Guest Comments of Online Hotel Distribution Sites. *Journal of Hospitality Marketing & Management*, (19) 7, pp. 773-796.
- Tambouratzis, G., Vassiliou, M., y Sofianopoulos, S. (2016). Language-Independent Hybrid MT: Comparative Evaluation of Translation Quality. *Hybrid Approaches to Machine Translation*, 131-157.
- Tan, Z., et al. (2020). Neural Machine Translation: A Review of Methods, Resources, and Tools. *AI Open*, 1, 5-21.
- Tertoolen, R. C. (2012). Desarrollos en la traducción automática: esperando aún una traducción de alta calidad. *Caracteres: estudios culturales y críticos de la esfera digital*, 1(2), 141-147.
- Toledo-Báez, M. C. (2018). Machine Translation and Post-Editing: Impact of Training and Directionality on Quality and Productivity. *Tradumàtica*, (16), 0024-34.
- Tomás-Gironés, J. (2003). *Traducción automática de textos entre lenguas similares utilizando métodos estadísticos*. (Tesis doctoral). Valencia: Universitat Politècnica de València. Recuperado de: <http://personales.upv.es/~jtomas/articulos/tesis.pdf>
- Torres-Hostench, O., Presas, M. y Cid-Leal, P. (coords.) (2016). *El uso de traducción automática y posesión en las empresas de servicios lingüísticos españolas: Informe de investigación ProjeTA 2015*. Bellaterra: <https://ddd.uab.cat/record/148361>
- Tripadvisor.com (2017). Worldwide Travel Path to Purchase 2017 (Analysis Based on April - September 2017 Data Periods).
- Tripathi, S., y Sarkhel, J. K. (2010). Approaches to Machine Translation. En: *Annals of Library and Information Studies* (57), pp. 388-393. Recuperado de: [https://www.researchgate.net/publication/228574546\\_Approaches\\_to\\_machine\\_translation](https://www.researchgate.net/publication/228574546_Approaches_to_machine_translation)
- Tucker, Jr. A. B. (1984). A Perspective on Machine Translation Theory and Practice. *Communications of the ACM*, 27(4), 322-329.
- Tucker, T. (2011). Online Word of Mouth: Characteristics of Yelp.com Reviews. *Elon Journal of Undergraduate Research in Communications*, 2(1), 37-42. Recuperado de: <http://www.elon.edu/docs/e-web/academics/communications/research/vol2no1/04tucker.pdf>
- Turing, A. M. (2012). Computing Machinery and Intelligence (1950). *The Essential Turing: the Ideas That Gave Birth to the Computer Age*, 433-464.

- Turovsky, B. (2016, Abril) Ten Years of Google Translate. Google Blog – The Keyword. <https://blog.google/products/translate/ten-years-of-google-translate/>
- Tyers, F. M. (2009). Rule-Based Augmentation of Training Data in Breton–French Statistical Machine Translation. En: *Proceedings of the 13th Annual Conference of the European Association of Machine Translation, EAMT09* (pp. 213-218).
- Tyers, F., y Donnelly, K. (2009). Apertium-cy: a Collaboratively-Developed Free RBMT System For Welsh To English. *The Prague Bulletin of Mathematical Linguistics*, 91, 57.
- Uszkoreit, H. et al. (2009). *Hybrid Machine Translation*. [Presentación en Power Point]. *TRANSLINGUAL EUROPE 2009*, Praga, República Checa.
- Valant, J. (2015). Online Consumer Reviews: The Case of Misleading or Fake Reviews. *European Parliamentary Research Service*.
- Valli, P. (2015). The TAUS Quality Dashboard. *Proceedings of the 37th Conference Translating and the Computer*, 127–136
- Van der Meer, J., Görög, A., Dzeguze, D., y Koot, D. (2017). TAUS Quality Dashboard White Paper. <https://www.taus.net/insights/reports/taus-quality-dashboard-white-paper>
- Van Slype, G. (1979). Critical Study of Methods for Evaluating the Quality of Machine Translation. *Prepared for the Commission of European Communities Directorate General Scientific and Technical Information and Information Management. Report BR, 19142*.
- Vargas-Sierra, C. y Ramírez-Polo, L. (2011). The Translator's Workstations Revisited: A New Paradigm of Translators, Technology and Translation. Session 4-Tools for Translators. In *Tralogy I. Métiers et technologies de la traduction: quelles convergences pour l'avenir?* (p 13).
- Vasconcellos, M. (1986). Post-Editing On-Screen: Machine Translation from Spanish into English. In *Proceedings of Translating and the Computer 8: A profession on the move*.
- Vásquez, C. (2011). Complaints Online: The Case of Tripadvisor. *Journal of Pragmatics*, 43(6), 1707-1717. Recuperado de: <https://www.sciencedirect.com/science/article/abs/pii/S037821661000384X>
- Vásquez, C. (2012). Narrativity And Involvement in Online Consumer Reviews: The Case of Tripadvisor. *Narrative Inquiry*, 22(1), 105-121. Recuperado de: <https://www.jbe-platform.com/content/journals/10.1075/ni.22.1.07vas>
- Vásquez, C. (2014). 'Usually Not One to Complain But...': Constructing Identities in User-Generated Online Reviews. En: C. Tagg & P. Seargeant (Eds.) *The Language of Social Media* (pp. 65-90). Palgrave Macmillan, London. Recuperado de: [https://link.springer.com/chapter/10.1057/9781137029317\\_4](https://link.springer.com/chapter/10.1057/9781137029317_4)

- Vásquez, C. (2015). "Don't Even Get Me Started...": Interactive Metadiscourse in Online Consumer Reviews. En: E. Darics (ed.). *Digital business discourse* (pp. 19-39). Palgrave Macmillan, London.
- Veale, T., y Way, A. (1997, February). Gaijin: A Bootstrapping, Template-Driven Approach to Example-Based MT. In *Proc. of the NeMNLP97*.
- von Helversen, B. et al. (2018). Influence of Consumer Reviews on Online Purchasing Decisions in Older and Younger Adults. *Decision Support Systems*, 113, 1-10. Recuperado de: <https://www.sciencedirect.com/science/article/pii/S0167923618300861>
- Wagner, E. (1985). Post-Editing SYSTRAN, a Challenge for Commission Translators. Recuperado de: <https://www.semanticscholar.org/paper/Post-editing-SYSTRAN%2C-a-challenge-for-commission-Wagner/672ef6e830ca6081d1e8891edee12a41f722ab64>
- Way, A., y Gough, N. (2005). Controlled Translation in an Example-Based Environment: What Do Automatic Evaluation Metrics Tell Us? *Machine Translation*, 19(1), 1-36.
- Weaver, W. (1955). Translation. *Machine Translation of Languages*, 14, 15-23.
- Wells, W. et al. (2000), Advertising Principles and Practice, Prentice-Hall, Upper Saddle River, NJ.
- What is Machine Translation? Rule Based Machine Translation vs. Statistical Machine Translation. (s.f.). Recuperado de: <https://www.systransoft.com/systran/translation-technology/what-is-machine-translation/>
- Williams, R. et al. (2008). The Importance of User-Generated Content: The Case of Hotels. *The TQM Journal*. 22, 2, 117 – 128. Recuperado de: [https://www.researchgate.net/publication/242345919\\_The\\_importance\\_of\\_user-generated\\_content\\_The\\_case\\_of\\_hotels](https://www.researchgate.net/publication/242345919_The_importance_of_user-generated_content_The_case_of_hotels)
- Wu, G. et al. (2010) Distortion as a Validation Criterion in the Identification of Suspicious Reviews. En: *1st Workshop on Social Media Analytics*, Washington, DC, (pp. 10-13). Recuperado de: <https://dl.acm.org/doi/abs/10.1145/1964858.1964860>
- Wu, Y. et al. (2016). Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation. *arXiv preprint arXiv:1609.08144*. Recuperado de: <https://arxiv.org/abs/1609.08144>
- Yunker, J. (2006). Statistical Machine Translation Gets Real: A Profile of Language Weaver. *Global by Design*. 12, 8-19.
- Zaretskaya, A. (2017). Machine Translation Post-Editing at TransPerfect—the 'Human' Side of the Process. *Tradumàtica tecnologies de la traducció*, (15), 116-123.
- Zehrer, A., Crotts, J. C., y Magnini, V. P. (2011). The Perceived Usefulness of Blog Postings: An Extension of the Expectancy-Disconfirmation Paradigm. *Tourism management*, 32(1),

106-113. Recuperado de:

<https://www.sciencedirect.com/science/article/abs/pii/S026151771000124X>