

LINGÜÍSTICA

ESTUDIOS DE LINGÜÍSTICA APLICADA VI

EDITAN

TERESA MOLÉS-CASES
ROSA CURRÁS MÓSTOLES
CARLOS PERIÑÁN-PASCUAL
FRANCESCA ROMERO FORTEZA



Estudios de lingüística aplicada VI

Editan

Teresa Molés-Cases

Carlos Perriñán-Pascual

Rosa Currás Móstoles

Francesca Romero Forteza

2023

Colección *UPV Scientia*

Para referenciar esta publicación utilice la siguiente cita: Molés-Cases, Teresa *et al.* (eds.) (2023). *Estudios de lingüística aplicada VI*. Valencia: Editorial Universitat Politècnica de València.

Edición científica:

Teresa Molés-Cases
Rosa Currás Móstoles
Carlos Perrián-Pascual
Francesca Romero Forteza

Comité científico:

Alonso Almeida, Francisco. *Universidad de Las Palmas de Gran Canaria*
Briz Gómez, Antonio. *Universitat de València*
Cantos Gómez, Pascual. *Universidad de Murcia*
Carrió Pastor, Marisa. *Universitat Politècnica de València*
Casas Gómez, Miguel. *Universidad de Cádiz*
Garofalo, Giovanni. *Università degli Studi di Bergamo*
González García, Francisco. *Universidad de Almería*
González García, Virginia. *Universitat de València*
Mairal Usón, Ricardo. *Universidad Nacional de Educación a Distancia*
Mitkov, Ruslan. *University of Wolverhampton*
Martín Arista, Javier. *Universidad de La Rioja*
Olmo Cazevieille, Françoise. *Universitat Politècnica de València*

Agradecimiento especial a Françoise Olmo Cazevieille y Marisa Carrió Pastor, ambas de la *Universitat Politècnica de València*

© De los textos: sus autores

© 2023, de la presente edición: Editorial Universitat Politècnica de València
www.lalibreria.upv.es / Ref.: 6720_01_01_01

ISBN: 978-84-1396-189-7
<https://doi.org/10.4995/SCLIN.2023.672001>



Estudios de lingüística aplicada VI/ Editorial Universitat Politècnica de València.
Se distribuye bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional.

RESUMEN

El volumen *Estudios de Lingüística Aplicada VI* presenta una serie de contribuciones recientes en el campo de la Lingüística Aplicada. Dado que encontramos variedad de temas en los nueve capítulos que incluye, los contenidos del volumen se han aglutinado en dos grandes bloques. El primer bloque versa sobre el análisis del discurso y el segundo bloque sobre la enseñanza y el aprendizaje de lenguas.

Palabras clave: análisis del discurso, semántica, multimodalidad, enseñanza y aprendizaje de lenguas, segundas lenguas.

OBJETIVOS

Este sexto volumen continúa dando a conocer aspectos de la investigación en Lingüística Aplicada. Entre los autores de las contribuciones encontramos tanto investigadores noveles centrados en su tesis doctoral, como investigadores más experimentados.

APORTACIÓN DE LA OBRA

El monográfico recoge una selección de contribuciones en inglés y en español sobre Lingüística Aplicada y, más concretamente, sobre investigaciones en torno al análisis del discurso (oral y escrito) y a la enseñanza y el aprendizaje de lenguas.

EDITORES

TERESA MOLÉS-CASES es profesora titular en el Departamento de Lingüística Aplicada de la Universitat Politècnica de València. Sus intereses de investigación incluyen estudios de traducción, lingüística de corpus, lingüística cognitiva y adquisición del lenguaje. Ha publicado el libro *La traducción de los eventos de movimiento en un corpus paralelo alemán-español de literatura infantil y juvenil* en Peter Lang y varios artículos en revistas internacionales (algunos ejemplos incluyen *Review of Cognitive Linguistics*, *Perspectives*, *Meta*, *Languages in Contrast*, *Lebende Sprachen*). Pertenece a los grupos de investigación COVALT (Corpus Valencià de Literatura Traduïda) y GALE (Grupo de Análisis de Lenguas de Especialidad). También ha trabajado como investigadora y docente en la Universitat Jaume I y en la Universität Leipzig (Alemania).

ROSA CURRÁS MÓSTOLES es profesora de Inglés con Fines Específicos en el Departamento de Lingüística Aplicada de la Universitat Politècnica de València. Es licenciada en Filología Inglesa por la Universitat de València (1985-1990) y doctora en Filología por la Universitat Politècnica de València (2009). Posee experiencia docente de inglés con fines específicos en diversas universidades tanto nacionales como internacionales. Su actividad investigadora se centra en la traducción cultural y la internacionalización de la educación superior. Ha participado como investigadora en varios proyectos de investigación, incluyendo el *Joint Research Centre* de la Unión Europea, y colaborado como evaluadora de proyectos europeos para el Servicio Español para la Internacionalización de la Educación Superior (SEPIE) del Ministerio de Universidades.

CARLOS PERIÑÁN-PASCUAL es doctor en Filología Inglesa y profesor titular en la Universitat Politècnica de València. Sus intereses investigadores se han centrado en estas últimas dos décadas en la ingeniería del conocimiento, el procesamiento del lenguaje natural, la lingüística computacional y la minería de textos. Más concretamente, es el responsable de la implementación computacional de una base de conocimiento para la comprensión automatizada del lenguaje, un sistema de extracción terminológica en dominios especializados, un entorno de experimentación en procesamiento de datos textuales y un programa para la detección automática de problemas publicados en las redes sociales, entre otras aplicaciones informáticas. Su producción científica incluye numerosas publicaciones en los campos de la lingüística, el procesamiento del lenguaje natural y la inteligencia artificial.

FRANCESCA ROMERO FORTEZA es doctora en Filología Catalana y profesora del Departamento de Lingüística Aplicada de la Universitat Politècnica de València, donde imparte asignaturas de Catalán para fines específicos en diversas titulaciones de ingeniería. Actualmente es directora de su departamento. Su investigación está centrada en el Aprendizaje de lenguas asistido por la tecnología y, más específicamente, en la modalidad del autoaprendizaje. Otros temas sobre los que versa su investigación son la enseñanza de lenguajes de especialidad y el análisis de errores lingüísticos. Por su tesis doctoral obtuvo el Premio extraordinario que otorga la Universidad y también ha obtenido el Premio a la excelencia docente que otorga el Consejo social de la Universitat Politècnica de València.

Índice

Prólogo	1
SECCIÓN I. ANÁLISIS DEL DISCURSO	
Capítulo 1. Contenido generado por el usuario procesado con traducción automática neuronal: localización y calidad de traducción automática <i>Carmen Rosa-Sorlózano</i>	5
Capítulo 2. Language resources for personality and emotion analysis: The case of MBTIEmo <i>Hongling Gou</i>	25
Capítulo 3. Metaphorical representation of smell: an analysis of perfume TV commercials under the lens of filmic multimodal metaphoricality <i>Lorena Bort Mir</i>	43
Capítulo 4. “I don’t think we need a million”: A study of the multimodal distribution of vagueness in spoken business discourse <i>Omra Malladi</i>	57
Capítulo 5. Forensic linguistics: Legal language for the resolution of crimes <i>Laura Vela Serrano</i>	71
SECCIÓN II. ENSEÑANZA Y APRENDIZAJE DE LENGUAS	
Capítulo 6. Explorando el aprendizaje sostenible de segundas lenguas a través del intercambio virtual <i>Oksana Polyakova Nesterenko y Ruzana Galstyan Sargsyan</i>	85
Capítulo 7. Using the Storytelling Power of Animations in EFL Classrooms: Students’ Perspective <i>Atrin Kavyanicherati</i>	105
Capítulo 8. Innovación pedagógica en el aula: una propuesta de aprendizaje basado en proyectos <i>Mingfang Zheng y Jingyue Zhao</i>	121
Capítulo 9. Una aproximación al perfil del profesorado rumano de ELE de secundaria superior: evidencias para el diseño de una guía de buenas prácticas <i>Paula Andreea Stinga</i>	143

Prólogo

Este volumen recoge una selección de las contribuciones presentadas en el decimocuarto Seminario de Investigación organizado por el Departamento de Lingüística Aplicada de la Universitat Politècnica de València, celebrado en mayo de 2023. Cuenta con 9 capítulos divididos en dos grandes bloques temáticos: análisis del discurso y enseñanza y aprendizaje de lenguas. Las investigaciones, desarrolladas tanto por investigadores noveles como por investigadores experimentados, constituyen una muestra de los intereses y tendencias de investigación en Lingüística Aplicada en la actualidad.

La primera sección, sobre el análisis del discurso, incluye cinco capítulos. El capítulo 1, “Contenido generado por el usuario procesado con traducción automática neuronal: localización y calidad de traducción automática”, se centra en el estudio de la traducción automática inglés-español de un corpus de reseñas de turismo y restauración y el posible impacto de esta automatización en los usuarios. El capítulo 2, “Language resources for personality and emotion analysis: The case of MBTIEmo”, presenta una serie de recursos lingüísticos enfocados al análisis de las emociones y la personalidad, entre los que destaca la herramienta MBTIEmo. Seguidamente, el capítulo 3, “Metaphorical representation of smell: an analysis of perfume TV commercials under the lens of filmic multimodal metaphoricality”, es una investigación interdisciplinar que ahonda en temas como la multimodalidad y la metáfora a partir del género de anuncios televisivos sobre marcas de perfume. A continuación, el capítulo 4, “‘I don’t think we need a million’: A study of the multimodal distribution of vagueness in spoken business discourse”, sigue desarrollando el concepto de modalidad a partir del estudio del concepto de vaguedad en un corpus de vídeos de entrevistas del ámbito empresarial. La sección concluye con el capítulo 5, “Forensic linguistics: Legal language for the resolution of crimes”, que presenta una introducción al campo interdisciplinar de la Lingüística forense.

Los capítulos de la segunda sección se agrupan bajo el paraguas de la enseñanza y el aprendizaje de lenguas. El capítulo 6, “Explorando el aprendizaje sostenible de segundas lenguas a través del intercambio virtual”, presenta una iniciativa formativa que persigue organizar e implementar la cooperación pedagógico-competencial aplicada al aprendizaje de inglés para fines específicos en entornos virtuales. El capítulo 7, “Using

the Storytelling Power of Animations in EFL Classrooms: Students' Perspective”, profundiza en las ventajas del uso de herramientas de storytelling en el aula de inglés como lengua extranjera. El capítulo 8, “Innovación pedagógica en el aula: una propuesta de aprendizaje basado en proyectos”, desarrolla una propuesta didáctica para el español como lengua extranjera (ELE) sobre formación para el empleo basada en el aprendizaje por proyectos. El monográfico termina con el capítulo 9, “Una aproximación al perfil del profesorado rumano de ELE de secundaria superior: evidencias para el diseño de una guía de buenas prácticas”, una contribución sobre cómo mejorar el aula de ELE en el sistema educativo rumano a partir de una investigación sobre la identificación de los principales rasgos del profesorado rumano de ELE de secundaria superior.

Esperamos que el volumen resulte de interés para investigadores y docentes en Lingüística Aplicada, sirva de inspiración para futuros estudios en este y otros campos afines y contribuya a un avance del conocimiento de carácter interdisciplinar.

SECCIÓN I

**ANÁLISIS
DEL DISCURSO**

Capítulo 1

Contenido generado por el usuario procesado con traducción automática neuronal: localización y calidad de traducción automática

Carmen Rosa-Sorlózano

Departamento de Lingüística Aplicada. Universitat Politècnica de València

1.1. Introducción

Los usuarios de internet han pasado a ser contribuidores activos en la Web 2.0. Pese al gran potencial y el volumen de negocio de este sector, las plataformas de reseñas traducen sus contenidos con traducción automática (TA) sin posterior revisión. Sin embargo, al ser un contenido con gran carga cultural, la posesición de estas reseñas debería considerar otros aspectos, además de la habitual corrección lingüística y estilística. Es decir, que se debería tener en cuenta la adaptación del mensaje a las preferencias lingüísticas y culturales específicas del usuario, algo que también se conoce como localización, que en los últimos 30 años ha contado con un papel clave en los Estudios de Traducción (Esselink, 2000: 11; Fernández-Costales, 2010: 25; Jiménez-Crespo, 2009: 3).

Por eso, este estudio se centra en un análisis de corpus de reseñas traducidas con TA. El objetivo de esta investigación es analizar el posible impacto de la TA sobre las opiniones de los usuarios, en concreto en el ámbito del turismo y de la restauración, para comprobar el grado de aceptabilidad del texto traducido con TA según los principios de la localización, de forma que se perciba el texto traducido como propio en el *locale* meta, es decir, en la cultura, el área geográfica o el mercado de destino del producto, para lo que a veces habrá que alejarse del texto original y adaptarlo en la medida necesaria.

Áreas como el turismo, las finanzas y el marketing podrían verse beneficiadas si mejoran sus sistemas de traducción, puesto que un mejor entendimiento de los servicios anunciados facilita la interacción del consumidor con dichos productos y servicios. La investigación permitirá extraer conclusiones sobre cómo afecta la TA sin poseer a la toma de decisiones de compra de los usuarios de internet y cómo la posesión y la localización de este tipo de textos es crucial dentro de la estrategia de marketing e internacionalización empresarial.

La importancia del estudio reside en la escasa bibliografía sobre esta temática. Existen artículos sobre traducción y turismo (Calvi, 2019; Cronin, 2000) o sobre TA y contenido generado por el usuario (CGU) (Candel-Mora, 2015), pero estos se han centrado en la terminología específica del turismo, las técnicas de traducción, los principales paradigmas de TA, estudios sobre los diferentes tipos de CGU y sus características, entre otros. Sin embargo, no se ha encontrado literatura que aúne todas estas variables. De esta manera, el presente estudio abarca todas estas temáticas con el fin de proporcionar mayor claridad sobre las posibilidades de futuro de los profesionales de la traducción.

Esta investigación ha consistido en el estudio de un corpus de 500 reseñas (250 de hoteles y 250 de restaurantes), extraídas de la plataforma de reseñas en línea TripAdvisor.com y escritas originalmente por hablantes nativos de inglés. A su vez, este corpus se ha dividido en reseñas de hoteles y restaurantes ya que estas presentan características diferentes entre sí, que determinarán cómo las procesan los motores de TA. Posteriormente, estas reseñas se procesaron con traducción automática neuronal (TAN) al español, para realizar estudios sobre la tipología de los errores de traducción y la calidad de los resultados de TAN. Finalmente, se ha llevado a cabo un análisis que ha permitido observar cómo la corrección o incorrección del CGU, así como su mayor o menor localización al *locale* o cultura de destino, podrían influir en el comportamiento de los usuarios a la hora de determinar si adquieren o no un producto o servicio.

1.1.1. Contenido generado por los usuarios

Las nuevas tecnologías han hecho posibles nuevas formas de comunicarse de manera escrita (Jones, Chik y Hafner, 2015), entre las que se encuentran los blogs o las redes sociales, gracias a las cuales, surge la posibilidad de que los usuarios relaten sus experiencias en internet, lo que ha dado lugar a un nuevo paradigma en la comunicación, el «boca-a-boca electrónico» o *electronic word-of-mouth* (eWOM) (Manap y Adzharudin, 2013).

El eWOM evoluciona gracias a las redes sociales, blogs y otras plataformas de internet y entre sus principales características se encuentran la escalabilidad y la rapidez de difusión de este tipo de contenidos (Dellarocas *et al.*, 2004).

Dada la creciente cantidad de contenido en internet en los últimos años, se podría considerar que el CGU es un género digital relativamente nuevo (Candel-Mora, 2015; Pollach, 2006; Vásquez, 2015). Según García de Torres (2010), en los primeros años del CGU digital, este aparece en grupos de noticias, foros, páginas web y chats. Más tarde, aparecieron plataformas en línea que ampliaron las posibilidades de creación del CGU, lo que evidenció el potencial de transformación de la Web 2.0 (García de Torres, 2010; Hidalgo-Alcázar, Sicilia y Ruiz de Maya, 2015). La Web 2.0 se podría definir como el conjunto de sitios web y plataformas en línea que permiten a los usuarios consumir y crear contenidos, además de compartir información en internet, por lo que estos se convierten en prosumidores, es decir, ya no solo leen y usan la información sobre los productos o servicios, sino que además crean contenido en redes (Muller *et al.*, 2011).

Según Vásquez (2014), el CGU es una de las características principales de las redes sociales y permite que cualquier persona con acceso a internet pueda dar su opinión y contar sus propias experiencias con los productos o servicios adquiridos. Hoy en día, gracias al acceso generalizado a las redes sociales, se ha creado un nuevo canal informativo para los usuarios, que consumen dicha información, pero que también la crean.

1.1.1.1. Definición y características

Las investigaciones sobre CGU y reseñas en internet son relativamente nuevas, lo que es evidente, ya que no existe una definición unánime para este tipo de contenidos (Candel-Mora, 2015). Existen diferentes nombres para estos: «online consumer reviews» (Vásquez, 2012) «user generated product reviews», «product reviews» o «user opinions» (Ricci y Wietsma, 2006), pero las denominaciones utilizadas en este estudio son «contenido generado por el usuario» (CGU) y «reseñas generadas por el usuario» (RGU). En este contexto, se entiende CGU como el concepto general que engloba todo tipo de contenidos generados por el usuario en internet y RGU como un tipo concreto de CGU.

El CGU se podría definir como aquellos contenidos redactados de manera pública por los usuarios de internet, en redes sociales o plataformas en línea y que no están editados por profesionales (Fernández-Castrillo, 2014; García de Torres, 2010; Hosteltur, 2018; Presi, Saridakis y Hartmans, 2014). Es importante recalcar que los usuarios que elaboran estos contenidos no son profesionales, sino que contribuyen voluntariamente con información o archivos multimedia para ayudar o entretener a otros usuarios (Fernández-Castrillo, 2014).

Según Vásquez (2012) el principal propósito de las reseñas es calificar, describir, y dar recomendaciones a otros usuarios a favor o en contra de un producto o servicio concreto. Por eso, se podría afirmar que el propósito principal de los consumidores cuando escriben las reseñas es compartir sus experiencias, y evaluar las fortalezas y debilidades de cierto producto o servicio. En concreto, el tipo de CGU en que se ha centrado este estudio han sido las reseñas publicadas en plataformas en línea de productos o servicios turísticos o de restauración.

Como se ha detallado anteriormente, es vital que las empresas proveedoras de productos y servicios de turismo y restauración lleven a cabo una buena gestión del CGU, ya que este las ayuda en sus campañas de marketing, para conseguir que las búsquedas en internet de los usuarios se conviertan en ventas. Además, las reseñas pueden ayudar a los usuarios a sentir una mayor seguridad al adquirir un producto o servicio intangible recomendado por otro usuario.

El estudio del CGU relacionado con la traducción de dichos contenidos tiene gran relevancia, ya que, según un artículo de la revista *Harvard Business Review* (Kelly, 2012), el 72,1 % de los consumidores pasa más tiempo en sitios web que están traducidos a su idioma que en otros que no lo están. Es más, un 72,4 % de dichos consumidores afirmó que probablemente compraría antes un producto cuya información en internet esté traducida a su lengua. Finalmente, según esta encuesta, el 56,2 % de los usuarios afirmó que no le importaría pagar más dinero, si esto significa que el producto está traducido a su idioma.

Por lo tanto, sería posible afirmar que el CGU es muy importante en el proceso de adquisición de un producto o servicio, sobre todo si se tiene 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; Schemmann, 2011). Esto sucede porque los consumidores no pueden probar los productos o los servicios de antemano, ya que son intangibles, por lo que, dichos usuarios buscan en internet reseñas o información que hayan proporcionado otros viajeros anteriormente. Por esto se podría afirmar que el CGU les ayuda en la toma de decisiones previa a la compra y minimiza la sensación de riesgo e incertidumbre (Presi, Saridakis y Hartmans, 2014; Schemmann, 2011).

Según un estudio realizado por Destination Analysts (2019), las aplicaciones más utilizadas por los usuarios para planear sus viajes fueron tanto páginas de tipo agencias de viajes digitales (Expedia, Hotels.com, etc.), como plataformas de reseñas (TripAdvisor, Yelp, etc.), por lo que las reseñas se han convertido en un elemento muy útil para los consumidores que quieren organizar su viaje o incluso para los que ya están de viaje, pero buscan recomendaciones.

Aunque el CGU sea un concepto relativamente nuevo, este se ha estudiado desde perspectivas diferentes. Por un lado, se han realizado estudios en el ámbito empresarial y del marketing, por otro, también se han estudiado los aspectos psicológicos del CGU, para entender por qué los usuarios elaboran estos contenidos y qué efecto tienen sobre

otros consumidores. No obstante, también sería importante entender los aspectos lingüísticos del CGU, algo que se ha hecho en esta investigación que ha tenido en cuenta estos tres puntos de vista.

1.1.2. Localización

La localización aparece como tal aproximadamente en los años 80, por lo que se considera una disciplina relativamente nueva, si se compara con la traducción. Esta aparece gracias al creciente uso de los ordenadores personales y de los nuevos *softwares*, que debían adaptarse a las necesidades de cada *locale* o cultura meta. Por otra parte, también se hizo más necesaria la traducción de estos programas para que fueran más entendibles y fáciles de usar (Esselink, 2003; Fernández-Costales, 2010; García-Nevarés, 2016; Pym, 2012).

En los años 90, con la aparición y el desarrollo de internet, los productos multimedia y la documentación que los acompaña se empiezan a distribuir a escala mundial (Esselink, 2000). Como resultado, aparecieron las primeras asociaciones de localización, como la desaparecida *Localisation Industry Standards Association* (LISA) (Esselink, 2000; Fernández-Costales, 2010) o la *Globalization and Localization Association* (GALA). Gracias a estas asociaciones y a la aparición de nuevos proveedores de servicios lingüísticos, se consolida la «industria de la localización», en que se conforman dos sectores principales muy diferenciados: los fabricantes y los distribuidores de *software* y *hardware* y, por otro lado, el sector lingüístico, es decir, el de los localizadores y los proveedores de servicios de localización (Esselink, 2000; Jiménez-Crespo, 2009).

Uno de los principales cambios que sufrió el proceso de localización fue que, anteriormente, los productos multimedia se localizaban cuando estaban finalizados, pero, más tarde, los productos se empezaron a plantear siguiendo el concepto del *simultaneous shipment* (*simship*), es decir, se comenzaron a comercializar todas las versiones de un producto de manera simultánea, algo que creó la necesidad de desarrollar herramientas que ayudaran a acelerar el proceso de localización sin comprometer la calidad (Esselink, 2000). Otro cambio importante fue el cambio de perfil de los traductores, que ahora, además de tener conocimientos lingüísticos, debían tener también conocimientos técnicos (Arevalillo-Doval, 2000; Esselink, 2002).

También en los años 90, la globalización de los negocios empezó a ser más relevante en el por entonces incipiente mundo de internet, lo que repercutió en la localización de los sitios web, ya que estos debieron adaptarse a las lenguas y las convenciones culturales de los diferentes *locales* (O'Hagan y Ashworth, 2002). Además, con la aparición de nuevos productos de *software* y contenidos en línea, el perfil de los usuarios cambió, por lo que se hizo aún más evidente la necesidad de localizar los productos para que estos fueran más accesibles (Fernández-Costales, 2010).

La localización es una disciplina relativamente nueva, por lo que no existe una definición universal de este concepto. Por eso, en este estudio, se ha proporcionado una definición a partir de diferentes ideas de autores consagrados en la materia. Para la gran mayoría

de autores, la localización consiste en adaptar lingüística y culturalmente productos informáticos, multimedia o sitios web, así como la documentación que los acompaña al mercado de destino, para que se puedan comercializar estos productos de forma que dicho *locale* perciba el producto como propio u original de su área geográfica (Esselink, 2000; Jiménez-Crespo, 2009). Es más, Jiménez-Crespo (2009) afirma que, si los textos localizados se asemejan a otros textos escritos originalmente en el *locale* meta, se considerará que estos son de gran calidad.

La localización cuenta con características que la diferencian de la traducción, entre ellas la corta vida de los productos, la diversidad de géneros y ámbitos, así como el lenguaje utilizado en cada uno de ellos, que será diferente en función de la situación comunicativa, y las distintas cualidades de los profesionales de la localización (Estellés, 2010). Otra característica notable de la localización es que esta pone el énfasis en el *locale* meta, por lo que se suele permitir a los localizadores modificar el texto tanto como sea necesario para adaptar el producto a las convenciones de la cultura de destino (Bernal-Merino, 2013).

Estellés (2010) y Schäler (2008) también afirman que los objetos de trabajo en localización no son solo textos, sino gráficos, audio, vídeo, animaciones o fragmentos de código, entre otros. Además, según el tipo de producto, los localizadores utilizarán una herramienta de localización u otra (Schäler, 2008). Además, la localización no solo se centra en los aspectos lingüísticos del texto, sino que también tiene en cuenta otras convenciones culturales de los *locales* de destino, como son los símbolos (monedas, unidades de medida, etc.), los iconos, los colores, así como otros efectos visuales (Anastasiou y Schäler, 2010; O'Hagan y Ashworth, 2002; Ressin, 2015). No tener en cuenta estos aspectos podría dar como resultado una traducción que no funcione en el *locale* de llegada (Jiménez-Crespo, 2008).

Como se ha podido observar, la localización va un paso más allá que la traducción, ya que esta se enfoca en aspectos más técnicos. Además, los lingüistas que se quieren dedicar a la localización requieren una formación especializada que les proporcione los conocimientos técnicos e informáticos necesarios. Ahora que se han expuesto las principales definiciones y características de la localización, en el siguiente apartado, se profundizará en la traducción automática. Para ello, se han proporcionado algunas de las definiciones que se han dado hasta la fecha, y se han estudiado las características de los principales tipos de motores de traducción automática.

1.1.3. Traducción automática

Para entender mejor la traducción automática, se han estudiado diferentes aspectos relacionados con esta, como las principales definiciones, tipos y orígenes, entre otros, recogidas en algunos de los estudios de investigación en TA que existen hasta la actualidad. A lo largo del tiempo, han surgido una gran variedad de modelos de TA, que van desde sus inicios como TA basada en reglas, hasta la traducción automática neuronal (TAN), entre otros innovadores modelos. El estudio de la TA es muy relevante

ya que, en una sociedad globalizada, la necesidad de traducción es cada vez mayor. Así, se hace imprescindible comprender el funcionamiento de los sistemas de TA, ya que esto permite diseñar una estrategia de posesición y localización que pueda afrontar los problemas lingüísticos que la TA todavía no es capaz de afrontar por sí sola.

La TA se ha estudiado desde puntos de vista muy diversos. Esta no solo se ha estudiado desde el punto de vista lingüístico, sino también desde el área de la informática, las matemáticas y otras disciplinas más técnicas. Por este motivo, existen diferentes definiciones para un mismo concepto, aunque con matices distintos, en función del campo de estudio desde el que se investigue. En este estudio preliminar, se ha elaborado una definición propia a partir de las distintas definiciones que se han aportado desde que hay constancia de los estudios en TA.

En sus inicios, se define la TA como un proceso de traducción llevado a cabo por ordenador, de un lenguaje natural a otro sin ayuda de un traductor humano, ni siquiera en el proceso de posesición, ya que el texto debería ser comprensible y útil para el lector (Informe ALPAC, 1966). Yunker (2006) afirma que los textos que se traducen con TA son aquellos que seguramente no hubieran sido traducidos por un traductor humano. Hutchins (2001) precisa que los investigadores se han centrado más en la TA de textos científico-técnicos, ya que, en esta, las diferencias culturales son menores y no se necesitan traducciones perfectas en cuanto a estilo y gramática, sino traducciones que transmitan la esencia del texto original y que ahorren tiempo y dinero (Hutchins, 2001, 2003).

Gaspari y Hutchins, (2007) definen los sistemas de TA en línea como los servicios a los que los usuarios pueden conectarse en internet para introducir los textos que deseen traducir. Hutchins (2001) puntualiza que la TA no siempre tiene que ser perfecta, ya que siempre hay diferentes traducciones para un mismo texto. Por eso, opina que no se debería dejar de lado el uso de la TA en ciertos contextos, ya que, en función de las características y requisitos de dicho texto, a veces puede llegar a ser más rentable utilizar TA frente a traducción humana.

En general, sí existe acuerdo entre los autores sobre qué es lo que representa la TA, no obstante, dado que a lo largo de los años han surgido motores de TA de diferentes tipos, se considera que la definición de TA se debería relacionar siempre con cada tipo concreto de TA. Sin embargo, dado el rápido paso al que evolucionan los diferentes modelos, todavía parece algo complicado definirla de manera universal.

1.1.3.1. TA Neuronal

Como se ha mencionado anteriormente, existen muchos motores diferentes, pero el utilizado en esta investigación ha sido el paradigma de la traducción automática neuronal, uno de los más utilizados hasta la fecha. En los últimos años, los sistemas de TAN han sido los que han ocupado a los investigadores, pero también a la industria, algo que, según Castilho *et al.* (2017), se debe en parte a los resultados positivos obtenidos por

las traducciones realizadas con TAN según las evaluaciones automáticas como BLEU. Además, según Wu *et al.* (2016), los modelos neuronales suplen las carencias de todos los sistemas anteriores.

La TAN se basa en redes neuronales artificiales (RNA) y aprendizaje automático (ML o *machine learning* en inglés). Según Casacuberta y Peris-Abril (2017), los sistemas de TAN son un tipo de modelo estadístico compuesto por unidades de procesamiento interconectadas. Los parámetros para estos modelos se estiman a partir de corpus paralelos mediante algoritmos de aprendizaje automático. En estos sistemas de TA, las palabras se representan con vectores y se utilizan redes neuronales recurrentes (RNN o *Recurrent Neural Networks* en inglés) para procesar las oraciones (Casacuberta y Peris-Abril, 2017).

Desde el punto de vista del aprendizaje automático, un sistema de TAN es una red neuronal que calcula la probabilidad de que una oración en lengua meta (LM) corresponda a una determinada oración en lengua origen (LO) (Castilho *et al.*, 2017; Kalchbrenner *et al.*, 2017; Luong, Pham y Manning, 2015). El sistema se entrena, mediante el uso de un corpus bilingüe alineado, sin necesidad de información lingüística externa (Castilho *et al.*, 2017; Jean, Cho, Memisevic y Bengio, 2014; Luong, Pham y Manning, 2015).

Una de las arquitecturas más conocidas de estos modelos es aquella formada por un codificador (*encoder*) y un decodificador (*decoder*) (Bahdanau, Cho y Bengio, 2014; Jean, Cho, Memisevic y Bengio, 2014). Este funciona a partir de una red neuronal que estima la probabilidad de que una traducción en LM corresponda a la oración en LO. Dicha red neuronal se compone por un codificador, que es la red para LO y codifica la secuencia en LO en una representación numérica (un vector), y un decodificador, que es la red para LM, la cual, mediante la representación generada por el codificador, da como resultado una oración traducida (Bahdanau, Cho y Bengio, 2014; Kalchbrenner *et al.*, 2017).

Con el tiempo, se han ido desarrollando sistemas neuronales más complejos y se han introducido otras estructuras como las redes neuronales convolucionales (CNN o *Convolutional Neural Networks* en inglés) o la traducción *zero-shot*. En las arquitecturas basadas en redes neuronales convolucionales, el decodificador crea representaciones de cada palabra teniendo en cuenta el conjunto de palabras a su alrededor. Por ejemplo, puede tener en cuenta unas dos o tres palabras a la izquierda o a la derecha de la palabra en cuestión (Forcada, 2017).

En los sistemas que utilizan redes neuronales convolucionales, la información de cada nivel se comparte en la red neuronal completa, así, cada filtro en la convolución cuenta con la visión completa de la oración original y, de esta, la red seleccionará algunas características (Pouget-Abadie, Bahdanau, Merriënboer, Cho y Bengio, 2014). Según Gehring *et al.* (2016), estos modelos cuentan con arquitecturas más simples y son más rápidos, puesto que las capas convolucionales posibilitan la fácil identificación de las relaciones entre elementos de una oración.

Por otro lado, la traducción automática *zero-shot* representa otro de los grandes avances en TA. Esta se basa en sistemas que pueden traducir entre pares de lenguas que no han visto durante su entrenamiento, todo dentro de un mismo modelo multilingüe (Al-shedivat y Parikh, 2019). Otros avances recientes han planteado la posibilidad de entrenar sistemas multilingües que traduzcan entre diversos pares de lenguas dentro de un único modelo (Siddhant *et al.*, 2019). Además, según Siddhant *et al.* (2019), también se ha mejorado la calidad de las traducciones automáticas entre lenguas con menos recursos.

1.1.4. Métricas de TAN

Para entender cómo se pueden mejorar los sistemas de TAN, es esencial evaluarlos siguiendo ciertos parámetros. No obstante, todavía no existe una definición única de calidad en que estén de acuerdo tanto el mundo académico como la industria (Castilho *et al.*, 2017). Para Doherty (2017), una de las principales problemáticas de las métricas de calidad existentes es la falta de estandarización, que podría deberse al hecho de que algunas de las métricas de calidad toman como referencia el texto original y otras el texto meta.

La calidad de las traducciones realizadas con TAN se puede evaluar con mecanismos automáticos, como BLEU, o manuales como el MQM (Multidimensional Quality Metrics) o el Dynamic Quality Framework (DQF) de TAUS (el más utilizado hoy en día) (Lommel, Uszkoreit y Burchardt, 2014). El modelo DQF de TAUS se basa en tres parámetros de evaluación (*utility*, *time* y *sentiment*) y, según el contenido del texto que se vaya a traducir, cada uno de estos tendrá un peso o relevancia distinto (Valli, 2015). Según O'Brien *et al.* (2011), para TAUS la «utilidad» (*utility*) es la funcionalidad del contenido, el «tiempo» (*time*) representa el tiempo en que se tardó en traducir el texto y el «sentimiento» (*sentiment*) es cómo los usuarios perciben la imagen de cierta marca o producto. Los principales tipos de errores contemplados en este modelo son la precisión (*accuracy*); la fluidez (*fluency*); la terminología (*terminology*); el estilo (*style*); el diseño (*design*); las convenciones de los *locales* (*locale conventions*), es decir, las reglas culturales propias de los usuarios meta, como los diferentes usos de símbolos o iconos, entre otros; o la veracidad (*verity*). Este modelo cuenta además con diferentes niveles de gravedad según el tipo de error.

En el apartado siguiente, se presenta un resumen de los resultados de los análisis llevados a cabo en base a las pautas de las métricas de calidad descritas en el modelo DQF de TAUS.

1.2. Resumen de resultados

Para este estudio, se ha llevado a cabo un análisis de las problemáticas de localización en el caso de las reseñas de hoteles y de restaurantes generadas por los usuarios, publicadas en plataformas digitales de reseñas en línea, en este caso, TripAdvisor.

com y traducidas mediante TAN. Para este análisis se han dividido los corpus en reseñas de hoteles y de restaurantes, ya que estas presentan características distintas entre sí que determinarán cómo los motores de TAN procesan dichas reseñas.

Para realizar este análisis, se han recopilado dos corpus bilingües, uno de hoteles y otro de restaurantes, con un total de 500 reseñas escritas en inglés (250 de restaurantes y 250 de hoteles). Estas fueron extraídas de la plataforma Tripadvisor.com y traducidas al español mediante el motor de TAN de la plataforma, en este caso, el Traductor de Google. A continuación, se han analizado los resultados de las traducciones automáticas de las reseñas, siguiendo las pautas habituales de posesición y las métricas de calidad, para identificar los principales tipos de errores de localización que afectan tanto a nivel lingüístico como cultural a la calidad de las reseñas traducidas con TAN que aparecen en la plataforma. El motivo de dividir el corpus en reseñas de hoteles y reseñas de restaurantes ha sido que cada una de estas reseñas tiene sus propias características únicas (longitud, mayor o menor carga cultural o estructura, entre otras). Dichas características determinarán cómo se procesarán las reseñas en los motores de TAN y, por consiguiente, también la calidad de los resultados de TAN.

El análisis ha consistido en un análisis cuantitativo y cualitativo de los dos corpus bilingües de reseñas traducidas al español con TAN. Para ello, se ha examinado el número de errores de localización en cada corpus y se ha evaluado cómo afectan estos a la calidad de las traducciones de las reseñas. En el análisis cualitativo, se han estudiado los problemas lingüísticos y culturales de localización, además de las causas frecuentes que afectan a la calidad de las reseñas traducidas con TAN, que en este caso son aquellas surgidas por problemas de explicación o duplicidad, problemas por la calidad del motor de TAN y problemas que se han producido por un error en el texto original. Además, se han comparado los resultados de los corpus bilingües de reseñas de restaurantes y de hoteles para observar las diferencias entre ambos.

En el caso de las reseñas de hoteles, los principales problemas han surgido porque la TAN no ha traducido correctamente los nombres de lugares turísticos, así como otras denominaciones pertenecientes a los barrios o a las calles de la ciudad. En el caso de los restaurantes, que eran de comida tradicional valenciana, se ha observado gran incidencia de nombres de ingredientes y platos regionales, lo que ha afectado negativamente a la TAN, que sigue teniendo problemas a la hora de traducir contenidos con altas cargas culturales. Tras aplicar diferentes criterios relacionados con la calidad de la TAN, estos errores se han clasificado para obtener unos resultados que ilustren cómo estos problemas afectan a la localización de las reseñas en cuestión. Entre otros, los principales errores cometidos por la TAN tanto en el corpus de reseñas de restaurantes como en el de hoteles son los que se muestran en la Tabla 1.1.

Tabla 1.1. Principales errores de TA. Corpus hoteles y restaurantes.

Tipo de error	Texto Original	TA ES	Comentario
Uso incorrecto del verbo «tener» cuando en inglés aparece <i>to have</i> en lo referente a comida.	<i>We had breakfast</i>	Teníamos un desayuno.	Uso incorrecto en ES de este verbo. Se podría haber traducido como «Tomamos un desayuno» o simplemente «Desayunamos».
Uso equivocado de «ser» y «estar» cuando en inglés aparece <i>to be</i> .	<i>The breakfast was standard, wasn't diverse enough in my opinion</i>	El desayuno era estándar, no estaba lo suficientemente diverso en mi opinión.	«El desayuno [...] no era lo suficientemente variado».
Traducción literal o no traducción de expresiones idiomáticas del inglés al español.	<i>Hop on hop off bus</i>	Autobús Hop on hop off	TA ha dejado la expresión sin traducir cuando en español si hay un término equivalente: «autobús turístico».
Uso de palabras de variantes del español de Latinoamérica, cuando en este caso la traducción era al español de España.	<i>Our block</i>	A una cuadra	«En nuestra manzana».
Uso incorrecto de las oraciones de relativo.	<i>Every time I have received an excellent reception.</i>	Cada vez que he recibido una excelente recepción	Lo que quiere decir el original es «Cada vez, he recibido una excelente bienvenida», y al introducir el «que» se modifica el sentido de la oración.
Uso incorrecto o abusivo de oraciones pasivas.	<i>Wasn't brought</i>	No fue traído	En español no es tan corriente el uso de pasivas. Su uso reiterado crea una sensación de poca naturalidad.

La TAN ha presentado problemas sobre todo a la hora de traducir vocabulario con fuertes cargas culturales como, por ejemplo, los platos o ingredientes regionales de la cocina valenciana. Asimismo, en el caso de las reseñas de hoteles, la TAN también ha cometido errores en la traducción de los nombres de lugares emblemáticos de la ciudad. En el caso de los términos con altas cargas culturales, los principales errores cometidos por la TAN han sido la traducción inconsistente de dichos términos o, directamente, el uso de un término incorrecto. Un claro ejemplo de este tipo de errores es el caso de la palabra *prawns*, que se ha traducido de forma inconsistente como «gambas», «camarones» o «carabineros», entre otras denominaciones, a veces incluso dentro de una misma reseña.

Otra clase de errores detectados frecuentemente en el corpus son los que se han denominado errores de explicación o duplicidad, que se producen cuando se mantiene en la traducción la explicación o la repetición de una palabra entre paréntesis que aparecía en la reseña original en inglés para explicar un término desconocido para los usuarios. Dicha explicación es innecesaria para la cultura meta, ya que sus miembros ya conocen el término o concepto del que se habla, por lo tanto, que la TAN haya mantenido estas explicaciones se considera erróneo. Además, también se han producido problemas originados a causa de la deficiente calidad del motor de TA o problemas derivados de errores en la reseña original. En las tablas siguientes (Tabla 1.2., Tabla 1.3. y Tabla 1.4.) se muestran algunos ejemplos de estos tipos de errores:

Tabla 1.2. Errores de explicación o duplicidad.

Original	TA ES	Comentario
<i>arroz al horno [...]</i> <i>(rice in the oven)</i>	arroz al horno [...] (arroz al horno),	Lo correcto en ambos casos hubiera sido eliminar el paréntesis.
<i>black pudding (morcilla)</i>	morcilla (morcilla)	

Tabla 1.3. Errores de calidad del motor de TA.

Original	TA ES	Comentario
<i>Massive bed</i>	Cama masiva	Cama enorme
<i>Fresh bathroom</i>	Baño fresco	Baño acabado de limpiar

Tabla 1.4. Problemas causados por errores en el original.

Original	TA ES	Comentario
<i>Quite tierd</i>	Bastante tierd	<i>Quite tired</i>
<i>You would kind</i>	Que te gustaría tener	<i>You would find</i>

En los análisis cuantitativos, se han identificado tanto el número de errores, como los principales tipos de errores encontrados en las reseñas y estos se han clasificado en errores lingüísticos o culturales. La Tabla 1.5. a continuación muestra algunos ejemplos de errores lingüísticos y culturales encontrados en los corpus de este estudio:

Tabla 1.5. Ejemplos de errores lingüísticos y culturales del corpus de estudio.

Tipo de error	Original	TA ES	Comentario
Cultural	<i>A soupy paella</i>	Paella espesa	Plato correctamente denominado «arroz meloso».
Cultural	<i>Main market</i>	Mercado principal	El Mercado Central
Lingüístico	<i>We had a break</i>	Tuvimos un descanso	Tomamos un descanso
Lingüístico	<i>To have valuable family time together</i>	Tener un valioso tiempo en familia juntos	Pasamos un valioso tiempo en familia

Seguidamente, se presentan los resultados obtenidos en el análisis cuantitativo con el propósito de identificar la cantidad de este tipo de errores en el corpus de restaurantes y en el de hoteles. La Tabla 1.6., a continuación, muestra los resultados correspondientes al número de errores lingüísticos y culturales del corpus de reseñas de restaurantes:

Tabla 1.6. Errores de localización culturales y lingüísticos. Corpus restaurantes.

Total de reseñas	250
Nº reseñas con errores culturales	44
Nº reseñas con errores lingüísticos	108
Total de errores de localización	152

El análisis del corpus de restaurantes ha dado como resultado un total de 152 errores de localización, 44 de estos errores han sido de tipo cultural y un total de 108 han sido errores de localización de tipo lingüístico. A su vez, la calidad de los resultados de TAN también se ve afectada por otros factores como los errores ya presentes en las reseñas originales, los errores de explicación o duplicidad o los errores que comete el motor de TA a causa de la calidad de este. Los resultados obtenidos tras analizar estos factores se muestran en la Tabla 1.7. a continuación:

Tabla 1.7. Errores que afectan a la calidad de 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 1.7. se muestran los tipos de problemas que afectan a la calidad de la TA de las reseñas del corpus de restaurantes. Como se observa, los problemas de explicación y duplicidad han sido los más frecuentes con un total de 217 casos. En segundo lugar, se han producido los errores a causa de la calidad del motor de TA, con un total de 172 casos. Por último, también han surgido problemas a causa de errores en las reseñas originales, con un total de 41 casos.

Por otro lado, la Tabla 1.8. muestra los resultados correspondientes al número de errores lingüísticos y culturales del corpus de reseñas de hoteles:

Tabla 1.8. Errores de localización culturales y lingüísticos. Corpus hoteles.

Total de reseñas	250
Nº reseñas con errores culturales	101
Nº reseñas con errores lingüísticos	209
Total de errores de localización	310

En este corpus de reseñas de hoteles se han producido un total de 310 errores de localización. De estos, 101 han sido errores culturales y, 209 han sido errores de tipo lingüístico. Adicionalmente, tras analizar los tipos de errores de explicación o duplicidad, de calidad del motor de TA y de errores en el original, se han obtenido los datos expuestos en la Tabla 1.9.:

Tabla 1.9. Errores que afectan a la calidad de TA. Corpus hoteles.

Causas que afectan a la calidad de la TA	Nº errores
Problemas explicación / duplicidad	14
Problemas calidad motor TA	194
Problemas error en el original	32

En la Tabla 1.9. se presentan los tipos de problemas que afectan a la calidad de la TA de las reseñas del corpus de hoteles. A diferencia del corpus de restaurantes, en el corpus de hoteles, los errores de explicación y duplicidad han sido los menos frecuentes con solo 14 casos. En este caso, los errores más frecuentes han sido los causados por la calidad del motor de TA, con un total de casos de 194. Finalmente, en el corpus de hoteles se han producido 32 casos de problemas a causa de errores en las reseñas originales.

Los resultados han mostrado que, en el corpus de reseñas de restaurantes, se han producido muchos menos errores que en el corpus de reseñas de hoteles, lo que ha sido en parte inesperado, dado que, en general, la carga cultural de las reseñas de restaurantes ha sido mucho más elevada por la cantidad de ingredientes regionales. No obstante, las reseñas de hoteles han sido en general mucho más extensas, lo que también ha contribuido a que se produzca un mayor número de errores.

1.3. Conclusiones

Con esta investigación se ha trabajado para comprobar si sería necesario implementar una estrategia de posesición que tuviera en cuenta los principios de la localización en las reseñas generadas por los usuarios publicadas en plataformas de turismo de internet, que se hayan traducido con TAN. De esta forma, uno de los propósitos del estudio fue evaluar la calidad de los resultados de TAN mediante una escala basada en las métricas de calidad de TAUS. Esta sección presenta las principales conclusiones de esta investigación.

Este estudio partía de diferentes hipótesis concretas. La primera era que el CGU puede llegar a afectar a la percepción del producto o servicio por parte de los usuarios. La segunda, era que, si no se localizan estos contenidos, los usuarios de destino no podrán hacer uso de esta información porque esta no estará traducida a su lengua ni adaptada a sus convenciones culturales, lo que podría afectar a la venta del producto o servicio. La tercera hipótesis ha sido que, si estos contenidos solo se traducen usando TAN sin una posterior posesición, podrían producirse errores de localización, tanto culturales como lingüísticos, que podrían afectar negativamente a la percepción del producto o servicio. Una cuarta hipótesis ha sido que, pese a que la TAN pueda representar una herramienta de apoyo en este contexto, sigue siendo necesario llevar a cabo una estrategia de posesición por parte de un lingüista profesional y nativo que revise los contenidos teniendo en cuenta los principios de la localización.

Tras realizar los análisis cuantitativos y cualitativos de los resultados de TAN de las reseñas, se han podido extraer varias conclusiones sobre los errores encontrados y el efecto de estos sobre la calidad de estos textos. Por lo general, se han producido más errores de tipo lingüístico que cultural tanto en el corpus de restaurantes como en el de hoteles. En el caso del corpus de reseñas de restaurantes, los errores más recurrentes han sido los de explicación o duplicidad, ya que aparecen muchos nombres de ingredientes o platos regionales que se han explicado o traducido entre paréntesis en la reseña original. Por otra parte, en el corpus de hoteles, los errores más frecuentes han sido los que se han producido por la calidad del propio motor de TAN.

Asimismo, gracias al análisis del corpus se ha podido observar que los errores de localización son más frecuentes en el corpus de restaurantes, ya que estas reseñas suelen contener más términos con alta carga cultural que las reseñas de hoteles. Otro problema importante de localización que se ha producido en ambos corpus ha sido el uso de palabras de variantes latinoamericanas del español, las cuales en este estudio se han considerado erróneas, ya que las reseñas estaban destinadas a un público de España y, por lo tanto, se ha basado la investigación en esta variante del idioma. En el caso de los errores de tipo lingüístico, uno de los principales errores ha sido la traducción inconsistente de verbos o términos. De este modo, se podría concluir que la TAN sigue cometiendo errores en la traducción de contenidos con cargas culturales altas (errores de localización), además, también sigue traduciendo de manera inconsistente o errónea determinadas palabras y expresiones idiomáticas.

**Para seguir leyendo, inicie el
proceso de compra, click aquí**