

ARTÍCULOS DE INVESTIGACIÓN

Las industrias de la lengua en la sociedad digital

Antonio Pareja Lora*

Departamento de Filología Moderna. Universidad de Alcalá (UAH). Alcalá de Henares/
Guadalajara, España

 <https://orcid.org/0000-0001-5804-4119>

Recibido: 06/12/2022 Revisado: 04/07/2023 Aceptado: 04/07/2023 Publicado: 21/09/2023

*Correspondencia

Correo electrónico: antonio.parejal@uah.es.

Cómo citar:

Pareja-Lora, A. Las industrias de la lengua en la sociedad digital. *Lemma: Revista de comunicación* 1(1). <https://doi.org/10.20511/lem.2023.v1n1.1862>

Resumen

En este artículo se hace un análisis del pasado reciente, el presente y el futuro de las industrias de la lengua, entendidas como “el sector de actividad dedicado al diseño, producción y comercialización de productos y servicios relacionados con el tratamiento de las lenguas” (LANGUNE, 2021). En primer lugar, se introducen este término y los grandes impulsores de su crecimiento y avance en nuestros días. En segundo lugar, se detallan los hitos más relevantes alcanzados hasta la fecha en el campo de las industrias de la lengua. En tercer lugar, se presentan, por un lado, la red de investigación europea LITHME y sus principales objetivos y, por otro lado, una clasificación, fruto de esta red, de las distintas profesiones que dan cuerpo a las industrias de la lengua hoy en día. En cuarto lugar, también se discute el futuro de las industrias de la lengua a través de sus retos actuales para, finalmente, concluir que, en general, las industrias de la lengua gozan de buena salud en nuestros días, aunque son grandes los desafíos que habrá de afrontar en los próximos años.

Palabras clave: Industria, lenguaje, tecnologías de la información y de la comunicación, traducción, enseñanza de lenguas

Abstract

This article analyses the recent past, the present and the future of the language industry, understood as “the sector of activity dedicated to the design, production and marketing of products and services related to language processing” (LANGUNE, 2021). Firstly, it introduces this term and the main drivers of its growth and progress today. Secondly, it details the most relevant milestones achieved to date in the field of language industries. Thirdly, it presents, on the one hand, the European research network LITHME and its main objectives and, on the other hand, a classification of the different professions that give shape to the language industry nowadays, resulting from this network. Fourthly, it also discusses the future of the language industry through its current challenges, and it finally concludes that, in general, the language industry is currently in good health, although it has major challenges to face in the coming years.

Keywords: Industry, language, information and communication technologies, translation, language teaching

1. Introducción

El término industrias de la lengua y el concepto que éste designa no son de reciente creación¹. Marcos Marín (1992) se refería a las industrias de la lengua con estas palabras:

esta presencia del ordenador en todos los campos de la vida diaria, [...] ha traído consigo la necesidad de desarrollar una serie de *productos lingüísticos comerciales*, como procesadores de textos o de palabras, [...] diccionarios, agendas, listines telefónicos informatizados, sistemas de interacción con los usuarios, [...] por no poner sino ejemplos de cada día. *La producción de estos bienes, de equipo o de consumo, [...] es la actividad propia de las industrias de la lengua* (p. 6; énfasis del autor de este artículo).

Asimismo, también Marcos Marín (1992), haciéndose eco de las palabras de Ángel Luis Gonzalo², define las industrias de la lengua así:

una intensa actividad lingüística e informática, en todo el dominio del español, que se recoge en los nombres de **industrias de la lengua**, o, con el más pretencioso y menos proclive a la broma, de «**ingeniería lingüística**» (p. 6; énfasis del autor de este artículo).

No es de extrañar que Ángel Luis Gonzalo, ingeniero de telecomunicaciones en primer término, se refiriera a las industrias de la lengua como una forma de ingeniería; pero no deja de sorprender que Francisco Marcos Marín (a la sazón Catedrático de Lingüística General en la Universidad Autónoma de Madrid³) sí lo hiciera. De hecho, las industrias de la lengua no se restringen únicamente al ámbito de la ingeniería lingüística, pues incluyen todas las formas de trabajo lingüístico y con las lenguas, tanto aquellas que hacen uso de máquinas u ordenadores como aquellas que se realizan de forma manual (aunque cada vez sean menos frecuentes, como la corrección de textos sobre papel). Así, más recientemente, se ha definido las industrias de la lengua como “el sector de actividad dedicado al diseño, producción y comercialización de productos y servicios relacionados con el tratamiento de las lenguas” (LANGUNE, 2021).

Según señalan García Delgado, Alonso y Jiménez (2010), los grandes impulsores económicos del desarrollo de las industrias de la lengua (en concreto, para las lenguas de comunicación internacionales, como el inglés y el español) son dos:

Primero, el avance del proceso de globalización económica, política y cultural, [...] en un mundo más interrelacionado informativamente y con más acentuados y extendidos procesos de ósmosis cultural (énfasis del autor de este artículo).

1 **Agradecimientos:** Este trabajo es fruto de la red de investigación asociada a “LITHME” (*‘Language In The Human-Machine Era’*, EU COST Action CA19102), financiada con fondos del Programa Marco Horizonte 2020 de la Unión Europea. También deseo agradecer a los grupos de investigación FITISPos (UAH) y ATLAS (UNED) su apoyo continuo en el desarrollo de mis investigaciones. Especial mención merecen también los doctores Maarit Koponen (*University of Eastern Finland*) y Rui-Manuel Sousa da Silva, que han contribuido enormemente al desarrollo de los trabajos del WG7 de LITHME presentados en el Apartado 3.2.

2 Director del Instituto para la promoción de las Tecnologías de la Lengua (INTLE), <https://www.linkedin.com/in/%C3%A1ngel-luis-gonzalo-p%C3%A9rez-b1342931/>.

3 <https://www.linkedin.com/in/fmm09/>.

Y segundo,

el **despliegue de la sociedad del conocimiento**, cuyo principal soporte —de lo que se sabe y de cómo se transmite lo que se sabe— es la lengua. La hegemonía de la red de redes lo ilustra inequívocamente; no en vano, la semilla, la clave fundacional de Internet es la idea del intercambio de información y de conocimiento (énfasis del autor de este artículo).

A estos dos factores, LANGUNE (2021) añade un tercero: “**las nuevas tecnologías**” o, equivalentemente, “el cambio marcado por la tecnología”, del cual el despliegue de la sociedad del conocimiento sería una consecuencia más o menos directa.

En esta sociedad del conocimiento, globalizada y tecnificada, las industrias de la lengua son claves para el acceso a “la información globalizada” o al “mercado común de la información”:

la UE [Unión Europea] apela a todos para poner recursos en el desarrollo de proyectos multilingües que eviten en un futuro inmediato la total dependencia respecto al inglés en el acceso a la información globalizada (Internet, ordenadores, tabletas, móviles, teleaprendizaje, telecompras...). Lo hace porque es consciente de que **las llamadas «tecnologías de la lengua» [son] imprescindibles para garantizar la ansiada accesibilidad global [...]**. Y en la UE hay casi treinta lenguas nacionales y otras muchas regionales para incorporar al proceso, y construir así el imprescindible «mercado común de la información» (Gonzalo, Guerra Romero, Sánchez-Ocaña y Mengual, 2016, p. XIII; énfasis del autor de este artículo).

Como puede verse, la Unión Europea (UE) confía en que las tecnologías de la lengua (la parte más técnica o ingenieril de las industrias de la lengua) ayuden a garantizar el acceso a la información de todos sus ciudadanos. Esto incluye no sólo a los hablantes de las lenguas más habladas de la UE, sino también de los de aquellas más minoritarias. Por ello, las industrias de la lengua están también llamadas a posibilitar la consecución de otros tres objetivos clave en una sociedad cada vez más globalizada: (i) el pleno desarrollo de los derechos lingüísticos (Pupavac, 2012); (ii) la eliminación de las barreras lingüísticas (Gonzalo *et al.*, 2016, p. XII); y, en su defecto, (iii) una igualdad de oportunidades que sea económicamente viable (*ibid.*) en el acceso a la información y el conocimiento de la población.

El principal cometido de este artículo es mostrar un resumen del devenir de las industrias de la lengua en las últimas tres décadas, y que las ha convertido en pieza clave y de uso cotidiano en la sociedad actual, además de adentrar al lector en su futuro más inmediato a través de los problemas a los que se enfrenta actualmente.

Así, se ha estructurado el resto del artículo de la siguiente forma. En primer lugar, en la Sección 2, se presentan los hitos más relevantes alcanzados hasta la fecha en el campo de las industrias de la lengua. En segundo lugar, la Sección 3 se centra en el presente de estas industrias, y presenta, por un lado, la red de investigación europea LITHME y sus principales objetivos (Apartado 3.1) y, por otro lado, una clasificación de las distintas profesiones que dan cuerpo a las industrias de la lengua hoy en día, elaborada en el contexto de dicha red (Apartado 3.2). Por su parte, la Sección

4 se adentra en el futuro de las industrias de la lengua a través de sus retos actuales. En la Sección 5 se detallan las conclusiones de este trabajo. Finalmente, se han incluido también secciones específicas para los agradecimientos y para las referencias bibliográficas asociadas a este artículo.

2. El pasado de las industrias de la lengua: hitos más relevantes

De nuevo de acuerdo con Marcos Marín (1992, p. 8), los grandes retos que debían afrontar las industrias de la lengua en la década de los 90 eran los siguientes:

El mercado de servicios en lengua natural informatizada, previsto por la Comisión⁴ para sus relaciones con las industrias y los usuarios, abarca estos campos: herramientas integradas de composición, bases de datos, herramientas de educación y perfeccionamiento, interfaces con sistemas humanos, gestión integrada de documentos, I/O lingüístico y codificación, recursos lingüísticos, traducción por ordenador, red gestionada de intérpretes, asesoramiento y control de calidad, lenguas reducidas, software e internacionalización de sistemas, tratamiento del habla, teleservicios y terminologías para el intercambio de información.

Como puede observarse, muchos de estos retos se han alcanzado con éxito. A continuación, se enumeran los ejemplos más representativos.

1. Para empezar, las funciones de redacción y respuesta inteligentes de la plataforma integrada de Documentos de Google⁵ son capaces de sugerirnos fórmulas y expresiones más o menos habituales de escritura en función del contexto (es decir, el texto que hemos recibido, escrito o estamos escribiendo); de igual manera, esta plataforma, además de tantos otros entornos de edición, sea de correo electrónico (Gmail, MS Outlook, etc.), sea de textos o presentaciones (v. g., MS Word o MS PowerPoint), e incluso nuestros propios teléfonos móviles (inteligentes) incluyen funciones para la corrección automática de errores ortográficos y hasta gramaticales.
2. Por lo que respecta a las bases de datos, en este caso lingüísticas, se han construido y poblado un gran número de lexicones computacionales con vocablos del léxico general, en su mayoría ya probados y evaluados con éxito, como WordNet⁶ (monolingüe – inglés), o BabelNet⁷ (multilingüe – 500 lenguas); pero también se han construido varias bases de datos terminológicas (con vocabularios especializados) de gran valor, muchas de ellas multilingües, como IATE⁸ (creada y mantenida por la Comisión Europea) o los diccionarios terminológicos del TERMCAT⁹, disponibles en línea.
3. En lo tocante a las herramientas de educación y entrenamiento, si nos centramos en la enseñanza de lenguas, son claro ejemplo de este campo las distintas aplicaciones para dispositivos móviles (*apps*) más usadas para el aprendizaje de lenguas, como Babbel¹⁰ o DuoLingo¹¹, además de otras más específicas, como BusinesApp, pensada para enseñar a preparar y realizar presentaciones efectivas en inglés (Pareja Lora, Calle Martínez y Pomposo Yanes, 2016).

4 El órgano de gobierno por excelencia de la Unión Europea.

5 <https://support.google.com/docs/answer/9643962>.

6 <https://wordnet.princeton.edu/>.

7 <https://babelnet.org/>.

8 Interactive Terminology for Europe, <https://iate.europa.eu/home>.

9 <https://www.termcat.cat/es/diccionaris-en-linia>.

10 <https://play.google.com/store/apps/details?id=com.babbel.mobile.android.en>; <https://es.babbel.com/>.

11 <https://es.duolingo.com/>; <https://play.google.com/store/apps/details?id=com.duolingo>.

4. En cuanto a las interfaces con sistemas humanos, aunque haya caído en el olvido o incluso se desconozca a fecha de hoy, ya hubo una primera ola de asistentes virtuales a finales del siglo pasado. Tal es el caso, por ejemplo, de los asistentes virtuales que se incorporaron en las versiones de MS Office de finales de los 90 (*e. g.*, Clipo o Clippy¹²). Su capacidad de comprensión del lenguaje y/o de acertar en sus respuestas era más que aceptable; pero, lamentablemente, consumían muchos recursos (*i. e.*, memoria), eran un tanto entrometidos y ralentizaban mucho el proceso de edición, por lo que fueron eliminados en versiones posteriores. No obstante, esta idea ha cobrado nuevo impulso recientemente con los *chatbots*, o asistentes virtuales en línea (como Isidra, en el caso de la Universidad de Alcalá¹³, o BILLIE, en el caso de IKEA¹⁴), gracias a una potencia de hardware y de computación mucho mayores que las entonces disponibles.
5. Acerca de los recursos lingüísticos, además de los anteriores y de los que siguen, podrían resaltarse todas las tecnologías, los módulos, las herramientas y subsistemas que sustentan y que se integran en las grandes aplicaciones de procesamiento del lenguaje natural (PLN), como los etiquetadores morfosintácticos, los analizadores sintácticos, o los anotadores semánticos desarrollados para un número considerable de lenguas. Probablemente, una de las librerías más completas y utilizadas en este sentido en el ámbito latino sea FreeLing (Padró, 2011), multilingüe, de libre acceso y de código abierto, y cuya demo¹⁵ ilustra con bastante claridad cuál es la función (y la posible utilidad) de los distintos tipos de anotadores y analizadores mencionados en este apartado.
6. ¿Y qué decir de la traducción automática (TA) y de los sistemas de ayuda a la traducción? Ya el propio Marcos Marín (1992, p. 8) habla del sistema SYSTRAN y del lanzamiento de otro proyecto, EUROTRA, “mucho más ambicioso, [...] que ha servido de catalizador de esfuerzos en muchos países, como España, que sin él habrían tardado mucho más tiempo en incorporarse a esta corriente científica y también social”. No obstante, estos sistemas no han trascendido el ámbito profesional, siendo bastante más conocidos y utilizados comúnmente los traductores de Google¹⁶ y de DeepL¹⁷; y aún está por ver qué impacto tiene fuera de la Comisión Europea del sistema sucesor de SYSTRAN y EUROTRA, al que se ha denominado eTranslation¹⁸.
7. Finalmente, casi con seguridad, otro de los campos de aplicación de las industrias de la lengua más conocidos en la sociedad actual, también gracias a las funciones incluidas en nuestros teléfonos móviles, es el propio de las tecnologías del habla. En efecto, cada vez más personas utilizan la voz para dar órdenes a su teléfono o dictar mensajes que luego envían por WhatsApp o por correo electrónico. Asimismo, estas mismas tecnologías permiten la lectura de textos en voz alta de forma automática, dando así asistencia, por ejemplo, a personas con problemas de visión. No obstante, los exponentes más avanzados de este campo de aplicación son los asistentes virtuales de los dispositivos iPhone (Siri¹⁹) y de Amazon (Alexa²⁰). Estos otros asistentes no

12 Clipo en la versión española, Clippy en la versión inglesa: <https://www.treecbits.com/2019/03/24/que-paso-con-clippy-el-asistente-de-microsoft-office/>; https://es.wikipedia.org/wiki/Ayudante_de_Office.

13 <https://portalcomunicacion.uah.es/diario-digital/actualidad/la-uah-estrena-asistente-virtual-para-su-pagina-web.html>.

14 <https://www.ikea.com/es/es/customer-service/contact-us/>.

15 <http://nlp.lsi.upc.edu/freeling/demo/demo.php>.

16 <https://translate.google.com/?hl=es>.

17 <https://www.deepl.com/es/translator>.

18 <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation>.

19 <https://www.apple.com/es/siri/>.

20 <https://developer.amazon.com/es-ES/alexa>.

dejan de ser unos asistentes virtuales más potentes y generales que los ya comentados más arriba, equipados adicionalmente con una interfaz de procesamiento y generación de mensajes de voz para una comunicación oral y más natural con las personas (preguntas y respuestas, búsquedas de información, envío y ejecución de instrucciones, etc.).

Curiosamente, entre los campos prioritarios de la Comisión Europea que menciona Marcos Marín (1992, p. 8) no se incluía aquél en el que se ha producido uno de los mayores y más importantes logros de la sociedad del conocimiento. Este logro notable produjo un enorme cambio en nuestra forma de entender el mundo y de acceder a los datos, la información y el conocimiento: el nacimiento de los buscadores de información en la web. En realidad, en 1992 era inconcebible el fenómeno que supondría la aparición de buscadores como el de Google. No en vano, los buscadores de Altavista (Broadley, 2020), Yahoo!²¹, o Lycos (Valera, 2019) aún estaban siendo desarrollados cuando se elaboró el informe al que alude este autor; y aún faltaban 6 años para que el buscador de Google fuera lanzado oficialmente desde su propio dominio en la web, en septiembre de 1998 (Rubal Thomsen, 2018). Sólo quienes usamos los sistemas previos al buscador de Google sabemos lo que supuso su aparición no ya para fines académicos (investigación, búsqueda de recursos educativos, etc.), sino tan sólo para buscar fuentes de información relevantes para las palabras claves de búsquedas más genéricas. Definitivamente, el buscador de Google cambió para siempre nuestras vidas (para bien).

3. El presente de las industrias de la lengua

En esta sección, por un lado (Apartado 3.1), se presenta la red de investigación europea LITHME, que se preocupa de las consecuencias y la influencia que ejercen la expansión y el auge de todas estas nuevas tecnologías sobre el lenguaje, el trabajo lingüístico y sus profesionales; y, por otro lado (Apartado 3.2), se comenta uno de los resultados preliminares de esta red, una clasificación tentativa de los distintos tipos de profesionales que trabajan en el campo de las industrias de la lengua.

3.1. La red de investigación europea LITHME

La red de investigación europea LITHME²² comenzó su andadura en noviembre de 2020, aún bajo los efectos de las restricciones de la COVID-19, pero también tras haber podido comprobar la gran ayuda que han supuesto las nuevas tecnologías para poder desarrollar nuestras labores cotidianas (i.e., trabajo, estudios u ocio) desde casa. No obstante, el proyecto ya se había gestado y enviado a evaluación con anterioridad a la pandemia causada por el coronavirus, y no deja de sorprender lo oportuno de su aceptación y lanzamiento por este motivo. En efecto, LITHME se preocupa de cómo afecta y afectará el uso generalizado de las nuevas tecnologías a

1. los distintos campos de las industrias de la lengua (v. g., la traducción);
2. la forma en la que las personas se identifican con un determinado idioma;
3. la propia estructura del lenguaje y las formas de interacción y comunicación social.

²¹ <https://search.yahoo.com/>.

²² LITHME: '*Language In The Human-Machine Era*', EU COST Action CA19102, <https://lithme.eu/>. LITHME consta de investigadores de todos los estados miembros de la UE, además de otros países fuera de la UE (tales como Suiza, Perú o Brasil).

Al mismo tiempo, LITHME pretende arrojar una luz sobre las implicaciones éticas de las tecnologías lingüísticas emergentes, tales como (a) la desigualdad en el acceso a las tecnologías (en el caso de las lenguas minoritarias o con menos recursos lingüísticos digitales), (b) los problemas de privacidad y seguridad, o (c) las nuevas formas de ciberengaño (por ejemplo, las noticias falsas o *fake news*) y ciberdelincuencia que posibilitan los avances tecnológicos.

En definitiva, LITHME busca equipar a los lingüistas y a otras partes interesadas para la era hombre-máquina. Así, sus objetivos pueden resumirse en:

- preparar la lingüística y sus subdisciplinas para los cambios que se avecinan;
- facilitar el diálogo a largo plazo entre lingüistas y desarrolladores de tecnología.

3.2. Los profesionales de las industrias de la lengua hoy en día

En este apartado se muestra y discute una clasificación preliminar (no necesariamente exhaustiva, pero sí sumamente representativa) de los distintos tipos de profesionales que actualmente ejercen sus labores en el campo de las industrias de la lengua. Esta clasificación, incluida en la Tabla 1, es el resultado de un trabajo coral de diversos investigadores de la red LITHME.

Tabla 1.

La clasificación de los profesionales de las industrias de la lengua (grupo de trabajo n.º 7 de LITHME)

Profesionales de las industrias de la lengua	Proveedores de servicios lingüísticos	Intérpretes	Intérpretes de lenguas de signos		
		Traductores, localizadores, transcreadores, poseedores	Especialistas en traducción multimedia	Audiodescriptores	
			Especialistas en traducción audiovisual	Subtituladores interlingües	
				Subtituladores intralingües	Editores de subtítulos cerrados
		Editores, revisores, correctores de textos			
		Directores de proyectos lingüísticos, asistentes de servicios lingüísticos			
		Terminólogos/terminógrafos, lexicógrafos			
		Traductólogos			
		Filólogos			
		Profesionales de la lingüística de corpus			
	Profesionales de I+D de las industrias de la lengua	Lingüistas computacionales	Desarrolladores de programas y módulos de PLN		
			Consultores de tecnologías del lenguaje		
		Psicolingüistas			
		Lingüistas clínicos, logopedas			
	Analistas lingüísticos	Analistas del discurso, lingüistas forenses, fonetistas			
		Profesores de idiomas			
		Profesores de traducción e interpretación			
	Educadores de las industrias de la lengua	Consultores lingüísticos			
		Redactores técnicos, redactores y especialistas en lenguaje claro			
		Productores y directores de contenidos, redactores publicitarios (<i>copywriters</i>)			
		Comunicadores	Periodistas, editores (v. g., en revistas o en editoriales)		
			Comunicadores corporativos o institucionales	Profesionales de CRM, gestores de comunidades y redes sociales	

Más concretamente, la clasificación de la Tabla 1 se ha construido dentro del grupo de trabajo número 7 (WG7²³, por sus siglas en inglés) de la red de investigación, que se centra en el trabajo lingüístico y sus profesionales (es decir, los profesionales de las industrias de la lengua). El WG7 está formado por profesionales, expertos e investigadores de los diversos campos propios de las industrias de la lengua. En dicha tabla, las profesiones se presentan por nivel de detalle, siendo los términos que aparecen a la izquierda los más generales o abstractos y aquellos más a la derecha los más específicos.

El primer borrador de esta clasificación fue elaborado por un equipo reducido de investigadores dentro del WG7, del que formaba parte el autor de este artículo. Este borrador fue posteriormente sometido a debate en el seno del grupo de trabajo, siendo comentado, corregido y aumentado por el resto de sus miembros.

Dadas (a) la amplia gama de profesionales y expertos en este ámbito que aglutina el grupo de trabajo, y (b) la metodología de trabajo colaborativo seguida en la elaboración de esta clasificación, puede decirse que es una muestra sumamente significativa (si no integral) de las distintas profesiones que sustentan las industrias de la lengua en la actualidad. Gran parte de ellas (traductor, intérprete, editor, revisor, profesor de idiomas, logopeda, etc.) son conocidas por cualquier lector lego en la materia; y otras tantas son conocidas sólo por audiencias más especializadas, como probablemente será el caso de los lectores de este trabajo (lexicógrafo, terminólogo, traductólogo, etc.). Todas ellas quedan aquí sin comentar por razones de espacio. Sí se discuten a continuación, no obstante, aquellas otras de reciente creación, emergentes, o con mayor potencial en la actualidad.

De esta manera, entre los proveedores de servicios lingüísticos cabe destacar, por una parte, a los audiodescriptores²⁴, los editores de subtítulos cerrados²⁵ y los intérpretes de lenguas de signos; y, por otra parte, a los poseedores, los transcreadores, y los especialistas en traducción multimedia. Por un lado, la labor de los tres primeros colectivos profesionales es especialmente importante en la sociedad actual, dada la especial atención que se está dedicando en nuestros días a la inclusión de personas con diversidad funcional, como aquellas con problemas de visión (público objetivo de los descriptores de audio) o de audición (público objetivo de los editores de subtítulos cerrados y de los intérpretes de lenguas de signos). Por otro lado, la labor de los otros tres colectivos es resaltable por motivos bien diversos. Para empezar, los servicios de posesición son el resultado de la aparición y el auge de los sistemas de traducción automática. De hecho, estos servicios se encargan de mejorar la calidad de los textos obtenidos mediante estos sistemas, por lo que son de relativamente reciente aparición dentro del mundo de los servicios de traducción. Por su parte, los transcreadores son también fruto de un fenómeno ya mencionado previamente, el de la globalización. Así, estos profesionales son un tipo particular de localizadores que trabajan fundamentalmente en el mundo de la publicidad y la mercadotecnia y producen traducciones que requieren una adaptación extrema para respetar todos los aspectos culturales y pragmáticos

23 LITHME's Working Group 7: Language work, language professionals; <https://lithme.eu/working-groups/wg7/>.

24 La **audiodescripción** es una "modalidad de traducción multimodal o intersemiótica en la que se traducen imágenes estáticas o dinámicas a una lengua, entendida como sistema de comunicación verbal, gestual o escrito, propio de una comunidad humana" (<https://tracce.ugr.es/lineas-de-investigacion/>).

25 Los **subtítulos cerrados** (*closed captions*, en inglés) son subtítulos que sólo pueden aparecer con el uso de un decodificador. El decodificador puede estar conectado a un televisor o ir incorporado de fábrica en el propio televisor (por defecto, a partir de julio de 1993). La información de los subtítulos se transmite en un área especial del espectro de la señal de televisión (la línea 21 del intervalo de supresión vertical) (<https://www.ncicap.org/captioning-terms>).

(sentido literal, figurado, implícito y/o deseado) del mensaje inicial, rompiendo las barreras del idioma y de la cultura tanto de la lengua (o variante) origen como de la lengua (o variante) meta. Finalmente, los especialistas en traducción multimedia, ya de larga trayectoria en el mundo editorial (en la traducción de cómics, por ejemplo), están cobrando cada vez más relevancia en la sociedad actual por su papel esencial en la traducción y localización de videojuegos (Gallo, 2020).

También son de destacar en este apartado ciertos tipos de profesionales de otras áreas de las industrias de la lengua, como son (1) los redactores y especialistas en lenguaje claro²⁶ y (2) los profesionales de CRM²⁷, o los gestores de comunidades y redes sociales. Los primeros acercan y facilitan el acceso de todas las personas a la información, en tanto que simplifican, por ejemplo, la lectura de textos administrativos y legales (Badía, 2017). Esto también allana el camino para una mayor inclusión e inserción de todos los sectores y sustratos de la ciudadanía en la sociedad actual, así como para una mayor participación ciudadana en los procesos de toma de decisiones de ámbito público, cada vez más en boga en nuestros días. Por todo ello, el lenguaje claro se ha convertido incluso en objeto de desarrollo de su propia serie de normas internacionales (ISO, 2021). Por lo que respecta a los segundos, son un grupo de profesionales claramente nuevo, creado o impulsado por el advenimiento de las redes sociales (eminentemente Twitter en este caso, pero también Facebook, Instagram, etc.). Estos profesionales son los nuevos responsables de la reputación de las empresas presentes en las redes sociales mediante sus propias cuentas corporativas. Básicamente, se ocupan de leer, moderar y responder los mensajes (a veces relativos a quejas o reclamaciones) que reciben de los usuarios o clientes de sus respectivas compañías.

4. El futuro próximo: retos actuales de las industrias de la lengua

Todos estos logros (Sección 2) y carreras profesionales (Sección 3) no están exentos de riesgos y retos, a los que las industrias de la lengua deberán hacer frente en un futuro más o menos inmediato. En lo que sigue se mencionan aquellos más acuciantes o relevantes con respecto a los aspectos más destacados previamente en este artículo. Para una mayor claridad, se hará de forma paralela a los logros reseñados en la Sección 2.

1. Así, en primer lugar, aunque las funciones de redacción (o respuesta) inteligente de los editores de texto y/o de correos son sumamente útiles, estas ciertamente cubren los niveles ortográfico (morfosintáctico) y gramatical (sintáctico), aunque se agotan en el nivel léxico-semántico, en este último caso gracias únicamente a las funciones de sinonimia y/o de búsqueda en diccionarios o tesauros que vienen integradas en los programas de edición habituales (v. g., MS Word). El reto en este sentido es incrementar la inteligencia y la ayuda que prestan estos programas en el nivel semántico y elevarlas también a los niveles del discurso y la pragmática. Por ejemplo, las personas que escriben en una segunda lengua (e. g., un investigador o un científico español que tiene que escribir un artículo en inglés) necesitan saber qué verbos

26 “Una comunicación está en lenguaje claro [en inglés, *plain language*] si la lengua, la estructura y el diseño son tan claros que el público al que está destinada puede encontrar fácilmente lo que necesita, comprende lo que encuentra y usa esa información” (<https://plainlanguagenetwork.org/plain-language/que-es-el-lenguaje-claro/>).

27 Las siglas CRM provienen del término inglés *customer relationship management*, que en español podría traducirse como **gestión o administración de la relación con el cliente**. En este artículo nos fijamos, no obstante, en su vertiente más actual, el **CRM social** o aquél que se centra y encausa a través de las redes sociales.

coocurren con mayor frecuencia, si ése es el caso, con el término que están utilizando en esa segunda lengua (es decir, asistencia semántica acerca de colocaciones en la lengua en la que escriben). Asimismo, estas mismas personas necesitan saber si todos los vocablos y expresiones que han incluido en su texto se corresponden con el registro formal propio de la comunicación escrita y/o científica, y que se les propongan alternativas en caso contrario. Asimismo, no todas las personas tienen la misma facilidad para la redacción de textos de todo tipo, ni siquiera en su propia lengua. Las personas con más dificultades a la hora de redactar ciertos tipos de textos (legales, administrativos, formales, etc.) se verían muy beneficiadas si existieran sistemas de ayuda para la escritura de textos de ese tipo particular, tanto en su propia lengua como en una segunda lengua, como ya lo hacen algunos programas en un cierto número de ámbitos especializados y/o para algunas lenguas (da Cunha Fanego, Montané March y Hysa, 2017²⁸; Sayers, Sousa-Silva, Höhn et al. (2021), pp. 47-48).

2. En segundo lugar, aunque es cierto que ya se ha construido un número importante de bases de datos lingüísticas, tanto léxicas como terminológicas, éstas incluyen en general las lenguas más mayoritarias y aquellas que ya están dotadas de más recursos lingüísticos y digitales (corpus, etc.), sumamente necesarios para poblar estas bases de datos. No obstante, las lenguas minoritarias y con menos recursos siguen estando en una clara desventaja e indefensión en este sentido (Sayers, Sousa-Silva, Höhn et al., 2021, p. 9). Las nuevas tecnologías no pueden servir para que las lenguas menos favorecidas lo sean aún más; muy al contrario, deben servir para ponerlas al nivel de las más mayoritarias y con más recursos, y no sólo para preservarlas y documentarlas adecuadamente, sino para revitalizarlas. Las lenguas son patrimonio de todos; cuando una de ellas muere o agoniza, estamos perdiendo parte de nuestra propia riqueza personal y comunitaria.
3. En tercer lugar, es de esperar que la enseñanza de lenguas se beneficie cada día más de los logros obtenidos en el conjunto de las industrias de la lengua. No obstante, pocos son los estudios y artículos científicos educativos que incorporan, por ejemplo, el uso de herramientas y sistemas de la lingüística computacional en las clases de idiomas, sea como recurso educativo o como herramienta de evaluación. En este sentido, son pioneros los trabajos de Harbusch, Cameran, y Härtel (2014) y Cotelli Kureth, Delorme Benites, Lehr, y Steele (2021); véanse también los sistemas de ICALL²⁹ mencionados en Sayers, Sousa-Silva, Höhn et al. (2021, pp. 37-38). Muy probablemente, otros similares se unirán a ellos pronto.
4. En cuarto lugar, de los interfaces con sistemas humanos y asistentes virtuales puede decirse lo mismo que de los sistemas de ayuda a la escritura del punto 1: es necesario elevar urgentemente su inteligencia semántica, discursiva y pragmática. No obstante, tal y como se menciona en el primer informe fruto del proyecto LITHME (Sayers, Sousa-Silva, Höhn *et al.*, 2021, pp. 24-26), probablemente el gran reto en esta línea es mucho más básico. Primeramente, habrá que encontrar una forma lógica y computable de representar de forma global y comprensiva el lenguaje y las lenguas; es decir, de representar el discurso y el significado, no sólo de las palabras, sino también de las oraciones y de toda una alocución completa. Después, habrá que

28 Véase el sistema ArText, de estos autores, disponible en línea en <http://sistema-artext.com/ambitos-especializados>.

29 ICALL: Aprendizaje de lenguas asistido por sistemas computacionales inteligentes; se usan las siglas inglesas (*Intelligent Computer-Assisted Language Learning*).

encontrar la manera de hacer inferencias certeras sobre asertos así representados, para poder determinar la validez y/o el valor semántico, discursivo y pragmático de un enunciado o de un fragmento de discurso dados (en su contexto específico).

5. En quinto lugar, acerca de los recursos lingüísticos en general sólo se puede repetir lo ya dicho en el punto 2: es sumamente urgente equiparar el número y el tipo de recursos lingüísticos digitales existentes para lenguas más mayoritarias y aquellos desarrollados para lenguas minoritarias o con menos recursos. Éste es el objetivo, por ejemplo, del reto OpenASR21, organizado por las importantes instituciones norteamericanas NIST e IARPA³⁰, y quizá el reto más importante en esta área.
6. En sexto lugar, en lo tocante a la TA, para empezar, cabe resaltar que gran parte de los avances producidos recientemente en las industrias de la lengua tienen su origen en la introducción de sistemas de inteligencia artificial (o, más concretamente, de aprendizaje automático) en sus procesos de negocio. Muchos de estos sistemas están basados en la tecnología de redes de neuronas, y los sistemas de TA actuales no son ajenos a esta tendencia.

Por ello, presumiblemente, el primer reto al que se enfrentarán los sistemas de TA y de PLN basados en redes de neuronas es su nivel de precisión, adaptabilidad e interoperabilidad. En efecto, todos los sistemas de PLN, sobre todo los basados en métodos estocásticos, suelen llevar asociada una tasa de error en sus resultados. Ahora bien, los sistemas basados en redes de neuronas, por oposición a los sistemas estadísticos basados en ejemplos o a los basados en reglas, se comportan además como sistemas de caja negra: son crípticos y mínimamente modificables. Esto impide su reajuste y dificulta su integración con otros sistemas para minimizar o corregir sus errores inherentes.

El segundo reto que habrán de afrontar estos sistemas está ligado a su obsolescencia y su mantenibilidad. En realidad, los conjuntos de entrenamiento (corpus lingüísticos) y el aprendizaje automático permiten la creación de modelos lingüísticos computacionales estáticos, que no se corresponden con los modelos lingüísticos que manejamos los humanos, mucho más dinámicos. El lenguaje evoluciona, y por ello, también los modelos lingüísticos deben hacerlo; en caso contrario, se quedarán obsoletos, y con el paso del tiempo sus respuestas se harán cada vez menos adecuadas o más forzadas (anticuadas, de hecho). Para evitarlo, habrá que encontrar la forma de hacer crecer los corpus de entrenamiento de forma dinámica, automática, periódica, controlada y adecuada (con datos representativos y que no generen ruido, sino conocimiento), y de conseguir que los modelos lingüísticos se reentrenen y evolucionen también de forma dinámica, automática y periódica, para dar respuesta a la evolución de las lenguas.

Un tercer reto que debe mencionarse en este apartado, y que afecta no tanto a las tecnologías lingüísticas, sino a los profesionales de las industrias de la lengua, es el de encontrar el punto de equilibrio entre sus recelos y el uso cotidiano de estas tecnologías en su trabajo; es decir, por un lado, que los traductores no conciban la TA ni como una amenaza ni como una pérdida de tiempo; pero, por otro lado, que su utilización tampoco suponga una reducción de la

³⁰ <https://sat.nist.gov/openasr21>; NIST: *National Institute of Standards and Technology*; IARPA: *Intelligence Advanced Research Projects Activity*.

calidad de los servicios de traducción. En la consecución de este objetivo no están exentos de responsabilidad los profesionales a cargo de la formación de los futuros traductores, que deberán dotarse a sí mismos y a su alumnado de criterios y buenas prácticas certeros para una traducción de calidad haciendo uso de sistemas de TA en sus trabajos y servicios (por ejemplo, conforme a la norma internacional ISO 18587:2017).

7. Finalmente, quizá los retos principales de las tecnologías del habla sean (i) alcanzar una expresión más natural, sobre todo de las emociones (González Lamaña y Pareja Lora, 2018); e (ii) incorporar en su proceso las mismas inteligencias reseñadas para los asistentes virtuales del punto 4; además de (iii) incrementar la cobertura de las lenguas (y variantes) para las que ya se han generado este tipo de sistemas.

En el tintero se quedan, por falta de espacio, otros retos esenciales de las industrias de la lengua, como la detección (semi-)automática de noticias falsas y bulos; o de casos de ciberacoso y lenguaje ofensivo, especialmente en redes sociales. Puede encontrarse más información acerca de estos y otros retos similares en Sayers, Sousa-Silva, Höhn et al. (2021) y Gonzalo (2016). Ambas fuentes contienen listas más completas de los retos a los que se enfrentan actualmente las industrias de la lengua (más general los primeros, y más específica del español los segundos).

5. Conclusiones

En este artículo se han analizado el pasado reciente, el presente y el futuro de las industrias de la lengua, que aquí se define, siguiendo a LANGUNE (2021), como “el sector de actividad dedicado al diseño, producción y comercialización de productos y servicios relacionados con el tratamiento de las lenguas”.

Con respecto a su pasado reciente, se ha dado cuenta de los hitos más relevantes alcanzados hasta la fecha en este campo, centrándose en siete aspectos clave: (1) las herramientas integradas de composición; (2) las bases de datos lingüísticas (léxicas y terminológicas); (3) las herramientas de educación y la enseñanza de lenguas; (4) los interfaces con sistemas humanos; (5) los recursos lingüísticos (en general); (6) la traducción por ordenador; y (7) el tratamiento del habla.

Por lo que respecta a su presente, se ha ofrecido información relevante acerca de la red de investigación europea LITHME y sus principales objetivos, por un lado, y, por otro, de una clasificación de las profesiones más relevantes de las industrias de la lengua en la actualidad que se ha realizado en el seno de esta red. Esta clasificación agrupa estas profesiones, en primer término, en torno a cinco categorías: (A) los proveedores de servicios lingüísticos; (B) los profesionales de I+D de las industrias de la lengua; (C) los analistas lingüísticos; (D) los educadores de las industrias de la lengua; y (E) los profesionales de las áreas restantes de las industrias de la lengua. Asimismo, se han introducido y discutido en detalle aquellas de más reciente creación, emergentes, o con un mayor potencial, como son los audiodescriptores, los editores de subtítulos cerrados, los intérpretes de lenguas de signos, los poseditores, los transcreadores, los especialistas en traducción multimedia, los redactores y especialistas en lenguaje claro, los profesionales de CRM, o los gestores de comunidades y redes sociales.

Finalmente, se han presentado los retos a los que habrán de hacer frente las industrias de la

lengua en un futuro más o menos inmediato, entre los que cabe destacar los siguientes: (i) la incorporación de una mayor inteligencia (semántica, discursiva y pragmática) en las funciones de redacción (o respuesta) inteligente de los editores de texto y/o de correos; (ii) la ampliación del rango de representación de las bases de datos lingüísticas y los recursos lingüísticos en general, para dar mayor cobertura a las lenguas minoritarias y con menos recursos; (iii) la mejora de la enseñanza de lenguas, incluyendo en ella de alguna forma el uso de herramientas y sistemas de la lingüística computacional, sea como recurso educativo, sea como herramienta de evaluación; (iv) la búsqueda y concreción de una forma lógica, computable, global y comprensiva del lenguaje, las lenguas y del significado, además de los algoritmos para poder hacer inferencias certeras sobre asertos representados de esta forma, de cara al perfeccionamiento de los interfaces con sistemas humanos, asistentes virtuales y las tecnologías del habla (que, en el caso de estas últimas, también deberán de progresar hacia una expresión más natural, sobre todo de las emociones); y (v) la solución de los problemas que presentan actualmente los sistemas de TA en particular y de PLN en general: su precisión, su adaptabilidad, su interoperabilidad, su posible obsolescencia y su mantenibilidad, además de las reticencias, miedos y objeciones existentes aún en nuestros días en el mundo de la traducción a este tipo de sistemas.

No obstante, cabe concluir que, en general, las industrias de la lengua gozan de buena salud en nuestros días, aunque son grandes los desafíos que habrán de afrontar en los próximos años. Sólo nos queda ponernos manos a la obra para afrontarlos con éxito.

REFERENCIAS

- Badía, J. (2017). *Sobre el lenguaje claro*. <https://lenguajeadministrativo.com/sobre-el-lenguaje-claro/>.
- Broadley, C. (2020). *AltaVista Search Engine History Lesson For Internet Nerds – Learn how one of the web’s biggest accidental success stories evolved*. <https://digital.com/altavista/>
- Cotelli Kureth, S., Delorme Benites, A., Lehr, C., y Steele, E. (2021, September). Machine translation in the language classroom: Swiss data. En *The Łódź-ZHAW Duo Colloquium on Translation and Meaning, online*. Winterthur, 2-3 September 2021. ZHAW Zurich University of Applied Sciences, School of Applied Linguistics, IUED Institute of Translation and Interpreting.
- da Cunha Fanego, I., Montané March, M. A. y Hysa, L. (2017). The arText prototype: An automatic system for writing specialized texts. En A. Martins, A. Peñas (eds), *EACL 2017. 15th Conference of the European Chapter of the Association for Computational Linguistics. Proceedings of the Software Demonstrations*, pp. 57-60. Valencia, Spain, 3-7 de abril de 2017. Stroudsburg (PA): ACL (*Association for Computational Linguistics*).
- Gallo, M. (2020). *La traducción de videojuegos. Un nuevo tipo de traducción multimodal*. GRIN Verlag.
- García Delgado, J. L., Alonso, J. A. y Jiménez, J. C. (2010). Capítulo 1: Valor económico del español: una síntesis. En L. García Delgado, J. A. Alonso Lera, J. C. Jiménez, M. E. Bombarelli

- (coords.), *El español, lengua global. La economía*. Madrid: Editorial Santillana / Instituto Cervantes. https://cvc.cervantes.es/lengua/espanol_economia/01.htm
- González Lamaña, A. y Pareja Lora, A. (2018). NaturTtSML—un esquema de anotación para la mejora de la naturalidad en los sistemas de síntesis de voz. *E-Aesla* 2018 (4), 375-390. <https://cvc.cervantes.es/lengua/eaesla/pdf/04/37.pdf>
- Gonzalo, Á. L., Guerra Romero, J., Sánchez-Ocaña, E. y Mengual, E. (2016). Prefacio. En Á. L. Gonzalo (coord.), *Tecnologías del lenguaje en España. Comunicación inteligente entre personas y máquinas*. Madrid: Editorial Ariel, S.A./Fundación Telefónica (pp. XI-XIV). <https://1library.co/document/zw35nrly-tecnologias-del-lenguaje-en-espana.html>
- Gonzalo, Á. L. (coord.) (2016). *Tecnologías del lenguaje en España. Comunicación inteligente entre personas y máquinas*. Madrid: Editorial Ariel, S.A./Fundación Telefónica. <https://1library.co/document/zw35nrly-tecnologias-del-lenguaje-en-espana.html>
- Harbusch, K., Cameran, C. J. y Härtel, J. (2014). Automatic Feedback Generation for Grammatical Errors in German as Second Language Focused on the Learner’s Personal Acquisition Level. En *Proceedings of EUROCALL 2014 - CALL Design: Principles and Practice*. Groninga, Países Bajos, Agosto de 2014.
- ISO: International Organization for Standardization (2017). ISO 18587:2017, Translation services — Post-editing of machine translation output — Requirements. <https://www.iso.org/standard/62970.html>
- ISO: International Organization for Standardization (2021). ISO/WD 24495-1, Plain language — Part 1: Governing principles and guidelines. <https://www.iso.org/standard/78907.html>
- LANGUNE (2021). *Industrias de la lengua*. San Sebastián: LANGUNE (Asociación de Industrias de la Lengua en el País Vasco). <https://www.langune.eus/es/presentacion/industrias-de-la-lengua>
- Marcos Marín, F. (1992). Industrias de la lengua. *Idiomas – todo sobre los idiomas*, 13 (noviembre-diciembre 1992), 6-11. <http://www.lllf.uam.es/~fmarcos/articulo/92IDInLen.pdf>
- Padró, L. (2011). Analizadores Multilingües en FreeLing. *Linguamatica*, 3(2), 13-20. <https://linguamatica.com/index.php/linguamatica/article/view/115>
- Pareja Lora, A., Calle Martínez, C., y Pomposo Yanes, L. (2016). Aprendiendo a hacer presentaciones efectivas en inglés con BusinessApp. *RIED. Revista Iberoamericana De Educación a Distancia*, 19(1), 41–61. <https://doi.org/10.5944/ried.19.1.14616>
- Pupavac, V. (2012). *Language Rights – From Free Speech to Linguistic Governance*. Londres: Palgrave Macmillan UK.
- Rubal Thomsen, R. (25 de septiembre de 2018). 20 años de Google: Google, una historia de éxito. *La Vanguardia*. <https://www.lavanguardia.com/vida/junior-report/20180920/451931192581/google-stanford-buscador-page-brine-20-anos-historia.html>

Sayers, D., Sousa-Silva, R., Höhn, S. *et al.* (2021). *The Dawn of the Human-Machine Era: A forecast of new and emerging language technologies*. Report for EU COST Action CA19102 'Language In The Human-Machine Era'. www.lithme.eu

Valera, M. (2019). *Qué fue de Lycos, el buscador que todavía sobrevive en la era Google*. M4C, Marketing 4 Commerce. <https://marketing4ecommerce.net/que-fue-de-lycos-el-buscador-que-aun-resiste-a-la-era-google/>