



Fronteras Literarias de la Inteligencia Artificial: Análisis Textual, Traducción Automática, y Generación de Texto en la Literatura Comparada.

Literary Frontiers of Intelligence Artificial: Textual Analysis, Machine Translation, and Text Generation in Comparative Literature.

DOI: 10.32870/argos.v11.n27.2.24a

Antonio Rodríguez Jiménez

Universidad Autónoma de Guadalajara (MÉXICO)

CE: arodriguezj15@gmail.com

ID ORCID: 0000-0003-4387-7649



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).

Recepción: 26/07/2023

Revisión: 22/08/2023

Aprobación: 09/10/2023

Resumen:

Después de muchos años de estudio arduo en el campo de la literatura comparada, se visualiza un presente y futuro prometedor en el estudio que se enfoca en comparar y analizar obras literarias de diferentes culturas, idiomas y períodos de tiempo para identificar similitudes, diferencias y patrones. Entra la inteligencia artificial con un fuerte impacto en campos esenciales como el análisis textual, la traducción automática y en la generación de texto, entre otras. El objetivo de este estudio es analizar estos campos tan útiles para la literatura comparada desde el ámbito de la Inteligencia Artificial. La metodología está basada en un análisis cualitativo de estudios textuales y experimentos expresados en bases valiosas de artículos y volúmenes. En resumen, la intersección entre la Literatura Comparada y la Inteligencia Artificial ofrece nuevas oportunidades para analizar, comparar y comprender obras literarias de una manera más profunda y sistemática.

Palabras clave: Fronteras literarias; Inteligencia Artificial; Traducción automática; Análisis Textual; Literatura comparada.



Abstract:

After many years of arduous study in the field of comparative literature, a promising present and future is envisioned in the study that focuses on comparing and analyzing literary works from different cultures, languages and time periods to identify similarities, differences and patterns. Artificial intelligence enters with a strong impact on essential fields such as textual analysis, automatic translation, and text generation, among others. The objective of this study is to analyze these fields that are so useful for comparative literature from the field of Artificial Intelligence. The methodology is based on a qualitative analysis of textual studies and experiments expressed in valuable bases of articles and volumes. In summary, the intersection between Comparative Literature and Artificial Intelligence offers new opportunities to analyze, compare and understand literary works in a deeper and more systematic way

Keywords: Literary frontiers; Artificial intelligence; Automatic translation; Textual Analysis; Comparative literature.

Introducción.

Los algoritmos de IA pueden analizar grandes cantidades de textos literarios para identificar temas recurrentes, estructuras narrativas y otros elementos literarios. Esto puede ayudar a los investigadores a realizar comparaciones sistemáticas entre obras literarias de diferentes tradiciones. García Serrano y Menta Garuz (2022, pág. 19) analizan dos investigaciones de diverso signo guiadas por la inteligencia artificial dentro del campo de las Humanidades Digitales. El primero es una investigación muy conocida y exitosa de dos lingüistas que resuelven un caso de atribución de autoría a través de la construcción de un corpus digital de 150 obras de 40 novelistas italianos. El segundo es la investigación llevada a cabo en el corpus digital DIMH (El Dibujante Ingeniero al servicio de la Monarquía Hispánica. Siglos XVI-XVIII), una evolución de la Colección de mapas, planos y dibujos del Archivo General de Simancas (siglos XVI-XVIII), cuyo objetivo fue desarrollar herramientas de soporte a tareas de anotación semántica, búsqueda de información, extracción de relaciones ocultas en los textos y visualización de los resultados para facilitar la investigación de los historiadores. Para delimitar estos saberes es necesario visitar otras disciplinas muy relacionadas con las HD como la inteligencia artificial (IA), el procesamiento del lenguaje natural (PLN)



y el análisis de datos o minería de textos (*text mining*, en inglés), entre otras. La formación en HD debe capacitar para actividades en las que el humanista digital sea capaz de hacer compatibles las necesidades de soporte informático de los humanistas con los recursos y herramientas informáticas existentes o susceptibles de diseño e implementación: la anotación de corpus literarios y textuales, la catalogación de documentos para repositorios virtuales, la creación de contenidos digitales, o la gestión de bibliotecas físicas y virtuales. (García Serrano y Menta Garuz, 2022, pág. 20).

Estos investigadores llegaron a la conclusión en sus estudios que en los dos casos se presentaron algunos de los problemas y tareas de investigación que pueden plantearse en las Humanidades Digitales, pero, bien es verdad que hay otras tareas específicas que ya forman parte de las HD, como son la digitalización y preservación del patrimonio (Literatura, documentalistas), la geolocalización (Historia y Arqueología), la minería de textos para la detección de noticias falsas (Periodismo), la detección de plagio, la visión artificial (Arqueología o caligrafías en textos antiguos), o bien tareas como la estandarización de sistemas de almacenamiento de los objetos de estudio, la navegación y búsqueda en repositorios específicos textuales o multimedia, el análisis léxico y semántico de textos, las aplicaciones con sistemas GIS, la catalogación web con buscadores facetados o en texto libre y otras aplicaciones de la web semántica. (García Serrano y Menta Garuz, 2022, pág. 36).

La IA ha avanzado en la traducción automática, lo que facilita la comparación de textos literarios en diferentes idiomas. Esto permite a los investigadores acceder a una gama más amplia de literatura comparada. Generación de texto: Los modelos de lenguaje basados en IA, como GPT-3, tienen la capacidad de generar texto literario. Esto plantea preguntas interesantes sobre la creatividad y la originalidad en la literatura, así como la posibilidad de que las máquinas puedan crear obras literarias comparables a las humanas. También es útil para el análisis de sentimiento y estilo, pues se puede utilizar para analizar el sentimiento y el estilo en la literatura comparada. Los algoritmos pueden identificar matices en la escritura y ayudar a los investigadores a comprender cómo los autores utilizan el lenguaje para transmitir emociones y temas.

Al contar cada vez con mayor volumen de datos a analizar y que gran parte de estos sea en formato texto, resulta muy dificultoso para las personas poder entender y aprovechar el valor que ofrecen. La



clasificación automática de textos consiste en asignar a un documento de texto una serie de clases utilizando técnicas de *Machine Learning* basado en su contenido y los temas que lo componen. La clasificación automática tiene importantes aplicaciones en la administración de contenido, la minería de opinión, el análisis de reseñas de un producto, filtros de SPAM y análisis de sentimiento en redes sociales. Rodríguez, Litman, Bolaños y Taboada (2019) manifestaron en el Simposio Internacional de Inteligencia Artificial de la Plata, que en su trabajo trataron de explicar paso a paso una estrategia híbrida entre aprendizaje no supervisado y clasificación automática de textos basada en obras clásicas y literatura técnica trabajando sobre textos no estructurados y seleccionando las técnicas apropiadas para llegar a una aplicación concreta. Luego de realizar evaluaciones con conjuntos de libros, los resultados obtenidos permitieron verificar que su enfoque era efectivo para la asignación automática de categorías a obras literarias.

Para Rodríguez, Litman, Bolaños y Taboada (2019, pág. 23), debido al creciente número de blogs, sitios y utilización de texto para comunicarse, la importancia de la clasificación automática de texto está tomando relevancia como nunca antes. La clasificación de textos puede ser automatizada, siempre que se hayan preprocesado y preparado los documentos con que se vaya a trabajar. Contar con datos de calidad, y grandes volúmenes, resulta crucial ya que las entradas que reciba el clasificador van a afectar la efectividad que tenga. En este caso, al contar una buena base de libros para entrenar genera un clasificador efectivo que puede discriminar entre distintos géneros y determinar, por ejemplo, si un documento es técnico, o no, y cuál es su contenido. Además, “durante la realización del análisis encontramos libros que componen sagas, sin definirlo previamente, e influencia entre autores solamente por las palabras que utilizan. Podríamos afirmar que el modelado y la ejecución fueron correctos al ver las clases generadas y cómo se corresponden con los nuevos documentos”, (2019, pág. 23).

El objetivo de este estudio es analizar estos campos tan útiles para la literatura comparada desde el ámbito de la Inteligencia Artificial. La metodología está basada en un análisis cualitativo de estudios textuales y experimentos expresados en bases valiosas de artículos y volúmenes.

Discusión y resultados



Hernández Mercedes (2002) escribió que en la era de la comunicación y de las nuevas tecnologías, la difusión de información y de ideas no debería tener obstáculos y, por lo tanto, la diversidad de lenguas existentes no debería constituir una barrera. Sin embargo, frente a la sólo aparente globalización lingüística, se hace patente la necesidad de aplicar el principio de localización. Al no existir una lingua franca, aunque muchos quieran creer lo contrario, surge con fuerza la necesidad de contar con instrumentos que permitan acceder a cualquier tipo de información, independientemente del idioma en que esté escrita o expresada. Es aquí donde la traducción automática (TA) se perfila como una de las claves para superar, al menos en parte, los obstáculos en la comunicación. Asegura la investigadora que el inglés, lengua nativa para sólo el 8% de la población mundial, es según Jean-François Degremont, sociólogo francés consejero de WorldMedia, lengua vehicular para el 20 % de la humanidad. Estudios recientes apuntan que, por ejemplo, la información en inglés depositada en Internet sobrepasa el 90% del total (Abaitua, I).

Hernández Mercedes (2002, pág. 1) manifestó en su estudio que las respuestas obtenidas tras pedir a un variado grupo de personas que definiesen qué significaba para ellas TA son indicativas a la hora de establecer lo poco que se sabe con respecto al tema y la conveniencia de aclarar algunos aspectos básicos: la traducción no la hace una persona sino el ordenador; es algo relacionado con internet; es traducir directamente, sin revisiones ni nada; son una bobada, no sirven para nada; es una manera muy moderna de traducir; yo digo una cosa y tú la traduces; es un tipo, creo, más evolucionado de traducción simultánea; no tengo ni idea; sí, sí. son unos programas que se instalan en el ordenador para traducir; obviamente, es un tipo de traducción, pero no sé de qué va; ¿es algo nuevo? Es, pues, “nuestro objetivo provocar la curiosidad del lector neófito hacia esta tecnología, ofrecer información con el fin de acabar con ideas erróneas o distorsionadas, aclarar una serie de conceptos y poner de relieve su valor como instrumento de ayuda”. (Hernández Mercedes, 2002, pág. 2)

El TA, en sentido estricto, se entiende como “el proceso por el cual una máquina traduce un texto de una lengua a otra, subdividiendo la sintaxis, identificando las partes del discurso, intentando resolver eventuales ambigüedades y, por último, traduciendo los componentes y la estructura en la lengua de destino” Sin embargo, tal denominación abarca actualmente numerosas y variadas herramientas



informáticas, aclara la autora. Añade Hernández (2002) que un sistema de TA que ofrezca resultados perfectos sin intervención humana, que traduzca sin error el discurso oral, que pueda mantener el estilo de cualquier documento y acceder a infinitos diccionarios, que resuelva todos los problemas..., no deja de ser, hasta el momento, más que una aspiración. A pesar de todo, se puede afirmar que el estado actual de la investigación y desarrollo es ya de interés operativo notable.

El nivel de los traductores automáticos basados en Inteligencia Artificial ha avanzado significativamente en los últimos años y continúa mejorando. Existen métodos como la traducción automática neural (NMT), que es una tecnología que ha impulsado mejoras significativas en la calidad de la traducción automática. Los modelos basados en redes neuronales artificiales han mejorado la fluidez y la coherencia de las traducciones, lo que ha llevado a una mayor precisión en la interpretación de contextos y matices. Hay también modelos de lenguaje pre entrenados, que han demostrado una capacidad sorprendente para la traducción automática. Igualmente, existe la traducción automática multilingüe, que aborda cómo los sistemas de IA han avanzado en la capacidad de traducir entre múltiples idiomas, lo que facilita la comunicación global. Y también es destacable la traducción específica de dominio, que consiste en traductores automáticos con IA, que también han mejorado en la traducción de lenguaje especializado en campos como la medicina, la tecnología y el derecho. Esto es crucial para la traducción precisa de textos técnicos y profesionales.

Igualmente, sobresale lo que se conoce como la traducción con contexto y sentimiento, consistente en algunos sistemas de traducción con IA, que han comenzado a incorporar análisis de sentimiento y contexto para ofrecer traducciones más adecuadas a la intención y el tono del texto original. Por último, existe lo que se denomina la integración en herramientas cotidianas, que son traductores automáticos con IA, que se han vuelto omnipresentes en aplicaciones y servicios en línea. Sin embargo, es importante destacar que, aunque los traductores automáticos con IA han avanzado considerablemente, todavía pueden cometer errores y no siempre son adecuados para traducciones altamente especializadas o literarias. La revisión humana sigue siendo esencial para garantizar la calidad y la precisión en muchas situaciones.



Según Casacubierta y Peris (2017), en la actualidad se producen cantidades ingentes de texto que deben ser traducidos a otras lenguas con cierta rapidez. Estas tareas de traducción pueden ser llevadas a cabo por profesionales, los cuales garantizan la calidad de las traducciones, pero el proceso es generalmente lento y de elevado coste. Esto ha obligado a la semi automatización de estos procesos mediante el uso de sistemas de traducción automática (TA o MT del inglés “machine translation”). Un ejemplo de la importancia de la TA reside en el volumen de palabras que genera la industria de la TA y que es de más de 100 mil millones de palabras diarias. La TA es, según la Wikipedia, “Un área de la lingüística computacional que investiga el uso de software para traducir texto o habla de un lenguaje natural a otro”. El diccionario Oxford en español la define como “traducción realizada por computadores o máquinas adecuadas para este fin”. En éstas y en cualquier otra definición, aparece de forma implícita o explícita el uso de los computadores y por tanto la necesidad de disponer de unos modelos precisos y de unos algoritmos eficientes. Todos estos elementos son de una utilidad inimaginable para la literatura comparada.

Beltrán y Rodríguez (2021) escriben que los cambios tecnológicos han permitido que el Lenguaje Natural (LN de ahora en adelante), en el que hablan los seres humanos cotidianamente se pueda incorporar en los computadores para la realización de diversas tareas que ayudan al ser humano a una mejor redacción o comprensión de un texto extenso, traducción y otras. Sin embargo, lo que estudian en su artículo estos investigadores es la potencial aplicación en el campo de la Ingeniería de Software de modelos como GPT-3, BERT con su arquitectura Transformer y otros, además de las aplicaciones derivadas que nacen de estas como los chatbots y su relación con procesos propios del ciclo de vida del software. En este contexto, se explican las generalidades del LN, del LN Computarizado, Inteligencia Artificial, Redes Neuronales para exponer los modelos sobresalientes en el campo del Procesamiento de LN, las aplicaciones y finalmente la cercanía con la Ingeniería de Software para evaluar las perspectivas de este tipo de innovaciones.

El GPT que se conoce hoy día significa *Generative Pre-Trained Transformer* y es en resumen un modelo de procesamiento del lenguaje generado por la compañía de investigación en el área de Inteligencia Artificial (o en inglés *Artificial Intelligence*) *OpenAI*. GPT se ha considerado para extracción de



relaciones de conceptos en un texto, traducciones o respuestas a preguntas. Utiliza la arquitectura basada en transformadores para consolidar Modelos del Lenguaje. Comprende el entrenamiento de un conjunto de datos en el cual, el modelo aprende las tareas como respuesta a preguntas, traducción automática o resumen de textos sin ninguna supervisión explícita. Beltrán y Rodríguez (2021, pág. 23) Igualmente, el GPT-3 es la tercera generación del modelo de PNL de la organización OpenIA, es autorregresivo y usa como su predecesor Aprendizaje Profundo con el fin de producir texto similar al del humano. No está disponible para el público en general. (Beltrán y Rodríguez, 2021, pág. 23) Estos sistemas también sirven de ayuda a la literatura comparada si se utilizan con la ética adecuada. Los autores mencionados concluyen en su estudio que la generación de productos de software está mediada por el lenguaje natural en la comunicación, a través de sus distintas etapas permitiendo de esta manera que los resultados finales sean dependientes de la comprensión y precisión en dicha comunicación. El procesamiento natural del lenguaje y sobre todo los modelos GPT han abierto notablemente la brecha de comunicación e interacción que tienen el ser humano y la máquina, permitiendo de cierto modo democratizar la tecnología. Es posible que GPT-3 haya establecido el punto de partida para que la interacción entre hombre y máquina cambie, permitiendo la generalización de software de una manera mucho más rápida y directa, proporcionado al usuario la posibilidad de representar sus necesidades de un producto de software a partir del lenguaje natural. A pesar de que GPT-3 se muestra prometedor, la limitación de uso restringe la posibilidad de generar aplicaciones centradas en las necesidades generales de la comunidad vinculada al desarrollo de software. (Beltrán y Rodríguez, 2021, págs. 32-33).

Referencias

- Abaitua, J. (2002). *Traducción automática: presente y futuro*, <http://www.foreignword.com> (versión española)
- Beltrán, N. C. y Rodríguez E. C. (2021). Procesamiento del lenguaje natural (PLN): GPT-3, y su aplicación en la Ingeniería de Software *Tecnol.Investig. Academia TIA*, 8(1), pp. 18-37.
- Casacubierta Nolla. F. y Peris Abril, A. (2017). Traducción automática neuronal. *Revista Tradumàtica*. (15) 66-74. <https://doi.org/10.5565/rev/tradumatica.203>



- García Serrano, A. y Menta Garuz A. (2022). La inteligencia artificial en las Humanidades Digitales: dos experiencias con corpus digitales. *Revista de Humanidades digitales*. (7) 19-39.
- Hernández Mercedes, P. (2002). En torno a la traducción automática. Instituto Cervantes. N° 2.
- Rodríguez, G. H., Litman, J., Bolaños, A., Taboada, G. H. (2019). Un enfoque híbrido para la clasificación automática de obras literarias. XX Simposio Argentino de Inteligencia Artificial (ASAI). Universidad Nacional de la Plata. 2-15