



PUNTO DE REFERENCIA: LA INTELIGENCIA ARTIFICIAL EN LA FILOLOGÍA

Ana Isabel Tsutsumi H.

Imagina que vives en España en el siglo XIV y que llega a tus manos el manuscrito del *Cantar de mio Cid*. Te interesa difundirlo —incluso si algunas frases te resultan misteriosas— porque es un texto relevante para la historia. Pasas largas horas en los claustros que habitas copiando el libro; sobre todo, te gustaría descifrar algunas palabras que no parecen del todo a gusto entre las líneas que las arropan. Fantaseas con viajar en el tiempo y encontrar a su autor para preguntarle qué quiso decir. A veces, por la mañana, después de pasar la noche en vela por estos misterios, modificas una palabra de ese manuscrito; quizá el copista que te precedió se equivocó. Eres un copista (un simple copista, pensarán), pero tu tiempo y esfuerzo, así como tu emergente curiosidad por los idiomas antiguos, alimentarán el nacimiento de la filología.

Entre los objetos de estudio de esta disciplina se encuentran los textos antiguos, aquellos que lograron sobrevivir al tiempo, conservados en distintos repositorios, como pueden ser archivos, bibliotecas o acervos. Contienen conocimientos e información de tiempos pasados y son la materia prima de filólogos, epigrafistas, paleógrafos e historiadores de todo el mundo. Sabemos que los textos más antiguos fueron escritos en arcilla o en cuero o tallados en piedra, lo que les permitió resistir el paso de los años. Estas características constituyen sus fortalezas, pero representan un reto a la hora de su interpretación.

Las herramientas digitales se han vuelto importantes para llevar a cabo investigaciones filológicas, no solo porque son auxiliares en la con-

servación de los documentos, pues evitan su degradación al digitalizarlos; también permiten la portabilidad de los mismos en imágenes de suma nitidez, de tal manera que resultan accesibles para cualquier investigador, no importa el lugar donde viva. Otro ejemplo del creciente impacto positivo de los medios digitales es el uso de las bases de datos de los acervos y las bibliotecas, pues el empleo de programas específicos de clasificación y gestión hace posible obtener información de manera eficiente. Por supuesto, hace ya varias décadas usamos textos digitales; se suele decir que todo está en la "nube", un servidor que tiene una gran capacidad de almacenamiento y que, junto con el desarrollo de programas de algoritmos que facilitan la recopilación y el análisis

de millones de datos, así como la transferencia de esa información y ese conocimiento, han propiciado que nuestra sociedad, a través de científicos y tecnólogos, logre avances en todas las áreas del conocimiento.

En la actualidad, la inteligencia artificial (IA) permite reconstruir textos dañados, hace sugerencias para completar la información y guía la interpretación de sus contenidos, tareas básicas para cualquier filólogo. En este texto abordo dos casos distintos de investigación filológica donde las IA juegan un papel relevante; además, expongo su instrumentalización en la fonología, que tiene que ver con el lenguaje oral, pero que podría replicarse para el caso de los textos escritos y, por ende, de la interpretación de estos.



Ocote, *Prompt tlacuilos*, 2023. Imagen generada con Midjourney

Un grupo de investigadores en el que intervienen matemáticos, programadores y filólogos ha implementado la herramienta BERT.

ITHACA

El primer ejemplo tiene que ver con el uso de Ithaca, una herramienta derivada de una red neuronal profunda (*deep learning*) que aprende, procesa y produce el lenguaje de manera artificial, apoyando la restauración textual; además, su análisis propone atribuciones geográficas y cronológicas a las inscripciones griegas antiguas.¹ Una red neuronal como el *deep learning* es una herramienta de la inteligencia artificial que se usa de manera efectiva en el procesamiento del lenguaje natural y la lingüística computacional; fue desarrollada de manera extensa en las últimas décadas.² En este proyecto, un grupo de investigadores de diferentes áreas, como la informática, la historia, la arqueología, la epigrafía y la filología, contribuyen con su experiencia y su conocimiento para completar el texto faltante de un escrito en piedra, datado alrededor de los años 500 a 450 a. C. Uno de los objetivos es el de proponer respuestas para interrogantes acerca del lugar y el momento histórico relacionados con el contenido del documento.

La investigación ha demostrado que la herramienta Ithaca coincide en sus resultados, en general, con los propuestos por epigrafistas especialistas en el tema; este grupo de investigadores señala que dicha herramienta requiere ajustes para ofrecer respuestas más precisas, pero que el camino está trazado. Ithaca utiliza, en su proceso, información y datos sobre el griego antiguo, datos de carácter histórico e inclusive literario, como lugares y fe-

chas; Ithaca integra esta información, es decir, “alimenta su conocimiento”, y a la vez procesa los datos, realiza inferencias, usa parámetros relevantes (en los que relaciona variables), para luego producir resultados. Estos aspectos son, por supuesto, relevantes para los estudios filológicos, pues permitirán corroborar las hipótesis de reconstrucción de los textos estudiados y, tal vez, hacer proyecciones sobre otros textos.

BERT

El segundo ejemplo considera el caso de manuscritos antiguos de textos griegos que se transmitieron a través de copistas a lo largo de la historia. Muchos de ellos inclusive registran escritos que estaban inscritos en piedra. Un grupo de investigadores en el que intervienen matemáticos, programadores y filólogos ha implementado la herramienta BERT para que utilice una de las bases de datos más grandes de griego antiguo e identifique y corrija errores no detectados hasta ahora y que fueron introducidos por los copistas en el proceso de la transmisión textual.³ Mucha de la bibliografía de la antigua Grecia ha llegado hasta nosotros gracias a los copistas, quienes transcribieron los manuscritos de textos filosóficos, literarios, geográficos e históricos. Los textos originales contenían borrones, letras faltantes o frases que no se entendían. En ocasiones, los copistas corregían faltas o agregaban anotaciones. La labor del filólogo radica en compren-

¹ Y Assael, T. Sommerschild, B. Shillingford *et al.*, “Restoring and attributing ancient texts using deep neural networks”, en *Nature*, 2022, núm. 603, pp. 280-283. Disponible en: <https://doi.org/10.1038/s41586-022-04448-z>

² Ruslan Mitkov (ed.), *The Oxford Handbook of Computational Linguistics* (2a ed.), Oxford Handbooks, Oxford, 2022. Disponible en: <https://doi.org/10.1093/oxfordhb/9780199573691.001.0001>

³ C. Cowen-Breen, C. Brooks, J. Haubold y B. Graziosi, “Logion: Machine Learning for Greek Philology”, 2023. Disponible en: <https://arxiv.org/pdf/2305.01099v1.pdf>

der todos esos elementos: cuál fue el sentido de los comentarios o cuáles fueron las correcciones u omisiones que se hicieron y que no pertenecían al texto original.

BERT es una inteligencia artificial basada en la metodología del aprendizaje automático, *machine learning*, desarrollada precisamente para la interpretación de lenguajes naturales. Una inteligencia artificial requiere forzosamente bases de datos robustas. Los investigadores del proyecto señalan que “se usaron alrededor de 6.4 millones de palabras del griego antiguo”, así como los textos escritos por Miguel Psellos, un prolífico escritor bizantino que vivió en el siglo XI en Constantinopla. De acuerdo con los autores de la investigación, los resultados ayudaron a detectar errores de los copistas, errores tipográficos de las versiones impresas de los textos y también errores ocurridos en la digitalización de los mismos. El proceso muestra mejoras una vez que se añaden las contribuciones de los filólogos, en particular, en lo que se refiere a la interpretación de los textos.

Un tercer ejemplo es una investigación relacionada con la morfofonología. Mediante una inteligencia artificial se utilizan patrones y relaciones subyacentes en distintas gramáticas; hasta ahora, se ha encontrado que 58 lenguas de nuestro planeta comparten los mismos patrones.⁴ Los hallazgos de esta investigación podrían determinar aspectos relevantes sobre la gramática e inclusive la semántica en la lengua hablada, además de las relaciones que existen, a nivel lingüístico, durante el aprendizaje del idioma en la infancia. Lo interesan-

te de este estudio, realizado con lenguas naturales, es que demuestra la validez del uso de las tecnologías de la inteligencia artificial en las diferentes áreas de la lingüística, y que estos métodos también podrían utilizarse para los estudios filológicos, es decir, en el análisis de los textos escritos antiguos, pues nos permitirían observar los procesos fonológicos, morfológicos y sintácticos históricos de la lengua. Así podríamos saber de qué forma una lengua se ha ido modificando. Por ejemplo, con una metodología similar a la de esta última investigación, los filólogos, junto con programadores y matemáticos, pueden empezar a estudiar el corpus de cientos de textos antiguos escritos en lenguas antiguas (por ejemplo, en latín) para obtener datos sobre los cambios lingüísticos a través del tiempo y los resultados se podrían comparar con los vastos estudios que existen sobre lingüística histórica.

La instrumentalización de diferentes inteligencias artificiales en los campos de la filología y la lingüística se encuentra todavía en un proceso inicial y contiene incertidumbres en los resultados; los propios investigadores sostienen lo imprescindible del apoyo de especialistas en filología para determinar la validez de los procesos de análisis. Es absolutamente necesario en temas como la interpretación de los textos, que requieren una comprensión extensa del contexto histórico, cultural y social en que los textos originales fueron escritos o en los que se transcribieron para su preservación. También será útil en el trabajo de ediciones críticas que se vienen realizando en distintas universidades. Elaborar la reconstrucción textual de las fuentes es tarea primordial en la filología, pues a partir de dicha labor se pueden realizar estudios de índole literaria, textual, antropológica o lingüística.

⁴ K. Ellis, A. Albright, A. Solar-Lezama, *et al.*, “Synthesizing theories of human language with Bayesian program induction”, *Nat Commun*, 2022, núm. 13, p. 5024. Disponible en: <https://doi.org/10.1038/s41467-022-32012-w>



Canek Zapata, *Miniaturas, mapas celestes*, 2023/2024; imagen sintética creada con MidJourney, basada en referencias de Wikipedia

En México se albergan acervos tanto de manuscritos como de impresos de diferentes periodos. Esos repositorios preservan escritos no solamente en español, sino en lenguas originarias como náhuatl, otomí, mixe, zapoteco y mixteco que, en su conjunto, representan una gran fuente de conocimiento de nuestra cultura y nuestra sociedad. Con toda seguridad, pronto seremos testigos del uso de inteligencias artificiales en proyectos relacionados con los archivos resguardados por diversas instituciones, como el archivo de lenguas indígenas albergado en El Colegio de México, o la base de textos CORDIAM en la Academia Mexicana de la Lengua, o en el IIMAS de la UNAM, donde se llevan a cabo las investigaciones sobre el lenguaje natural. Si utilizamos las inteligencias artificiales en el ámbito de proyectos multidisciplinarios, obtendremos avances en la investigación filológica.

A la luz de las aplicaciones de la inteligencia artificial en la filología cuesta trabajo no pensar de nuevo en nuestro copista anónimo, que trabajó horas y días para ayudar a que el *Cantar de mio Cid* no solo no se perdiera, sino que fuera difundido y leído por más personas. La historia cambia o se complementa. Ahora es la filóloga que ayudó a diseñar BERT la que quisiera viajar en el tiempo. Le diría al copista del siglo XIV que no andaba tan errado al modificar esas palabras. Si el viaje en el tiempo existiera ya, esta filóloga viajaría con un documento que propone las palabras que cree correctas para el texto y así nuestro copista dormiría a pierna suelta. Pero el viaje en el tiempo aún no es posible y los especialistas de las distintas épocas tenemos que conformarnos con un diálogo entrecortado a través de los siglos, pero un diálogo, al fin y al cabo, que preserve viva la cultura. **U**



Santiago Moyao, *La galería*, 2023. Imagen generada con Midjourney