

3.1 Algoritmos y análisis de textos: una relación fructífera

El reconocimiento de patrones lingüísticos por medios informáticos revolucionó los estudios históricos y filológicos **mucho antes de la generalización de los LLM**. Un caso célebre del uso de herramientas digitales es el de la búsqueda de **la identidad de la escritora italiana Elena Ferrante**: en el año 2017, un grupo de expertos de la universidad de Padua analizó ciento cincuenta novelas de autores italianos contemporáneos para **tratar de identificar quién se escondía detrás del seudónimo** con el que se había publicado la famosa saga de novelas napolitanas que incluye *La amiga estupenda* y que se convirtió en una exitosa serie televisiva de la plataforma HBO. **El misterio sigue sin resolverse, así que aún no podemos saber si el análisis fue acertado.**

Pero el uso de ordenadores en la disciplina que se conoce como *estilometría* **ya se había utilizado desde la década de 1960**, cuando se aplicó por primera vez al Nuevo Testamento: el escocés Andrew Q Morton utilizó un ordenador de la universidad de Edimburgo para **llegar a la conclusión de que san Pablo solo había escrito cuatro de las catorce epístolas que se le atribuían**. La búsqueda de la presencia o ausencia de **estilemas** (rasgos estilísticos específicos de un escritor o escritora) se venía utilizando desde el siglo XIX para confirmar o rechazar la atribución de una obra a Shakespeare o a Lope de Vega, **pero las posibilidades se multiplicaron con el uso de bases de datos y herramientas informáticas creadas con ese propósito**. Además de atribuir autorías, la herramienta se utiliza también para establecer, por ejemplo, **la cronología de las obras de un autor**. En otras disciplinas humanísticas, los estudios del léxico y la sintaxis también se han utilizado, por ejemplo, para **determinar si dos culturas históricas determinadas tuvieron contacto en algún momento** (gracias al estudio de los préstamos lingüísticos), o el orden de redacción de las obras de Platón y de Aristóteles (un análisis que permite escrutar la **evolución de su pensamiento**).

El estudio estadístico del lenguaje, sin embargo, no solo se había utilizado para desentrañar oscuros asuntos académicos, y sus aplicaciones ya se habían generalizado, muchos años antes de que apareciera la IA, en campos tan variados como **el espionaje, la investigación de crímenes en serie, las previsiones de resultados electorales y los estudios de mercado**.

Una primera aplicación de este tipo de herramientas estadísticas aplicadas a textos de distinto tipo, tal vez la más evidente, es la **extracción de información** buscando coincidencias en uno o

varios documentos, o en internet. Este proceso básico, **similar a lo que hacemos cuando buscamos una palabra concreta en un texto** (o en un motor de búsqueda), se ha desarrollado en una tarea conocida como **NER** (por las siglas de su nombre en inglés, Named Entity Recognition) y se puede utilizar, por ejemplo, **para extraer nombres de empresas y personas de artículos de prensa, para analizar contratos e identificar a las partes firmantes, o para detectar menciones a empresas en distintos medios de comunicación.**

Otros usos previos a la generalización de la IA, permitían la **clasificación automática de textos** (por ejemplo detectar y separar el *spam*, como hacen los servicios de correo electrónico, o clasificar incidencias de clientes por tipo de problema).

Este tipo de herramientas se utilizaba también en universidades y editoriales para detectar **plagios o parecidos no razonables**, aunque no siempre con éxito.

Una vez más, la aparición de la inteligencia artificial generativa, y en concreto de los LLM, ha permitido que los análisis, reservados hasta ahora a los expertos con conocimientos especializados, estén al alcance de cualquier curioso que sea **capaz de crear un prompt adecuado en lenguaje natural**. Ya hemos visto que el desarrollo de los modelos de lenguaje se ha utilizado para realizar **análisis de sentimientos** (detectar si un texto expresa una opinión positiva, negativa o neutral). De hecho, muchas plataformas ya incorporan sus propios sistemas de IA que condensan las opiniones de sus clientes, tal y como hace Amazon y muchas otras empresas.

Con la IA, además, se han abierto otros campos que la tecnología anterior no era capaz de realizar. Algunas de las más habituales son las siguientes:

- **Resumen automático de textos** (por ejemplo informes empresariales, novelas completas, artículos científicos o incluso reuniones y entrevistas): algunas herramientas habituales son ChatGPT, de uso general, Notion AI (para empresas) y, en el caso del ámbito académico, NoteBookLM.
- **Detección de temas** (por ejemplo analizar miles de noticias para ver qué temas , examinar foros para detectar problemas recurrentes y analizar publicaciones científicas para ver tendencias de investigación. (tendencias de Google)
- **Análisis lingüístico avanzado** que permite reconocer la complejidad del lenguaje de un libro, analizar el estilo de una autora o autor o detectar discursos de odio o y mensajes tóxicos en redes sociales.

Si pedimos a un *chatbot* un **ejemplo de uso real**, encontramos la siguientes propuesta:

“ □ Ejemplo completo de aplicación real:



Un periódico analiza 100.000 tweets sobre una elección en Spain:

1. La IA clasifica los tweets por tema.
2. Analiza sentimiento.
3. Extrae nombres de políticos.
4. Genera resúmenes de tendencias.

Resultado: un informe sobre la conversación pública en redes.

Como se puede apreciar, el ejemplo es útil, pero también poco específico, y contiene errores e imprecisiones, además de algunas **expresiones poco afortunadas desde un punto de vista estilístico**. La aplicación real en unas elecciones en "Spain", en principio, solo puede llevarse a cabo si hay alguien detrás (un humano o grupo de humanos) que afine el *prompt* y sepa interpretar los resultados.

Las herramientas de IA de las que disponemos todavía no son capaces de realizar un análisis pertinente y con sentido de ningún texto, **salvo que se les ofrezcan unas coordenadas con las que trabajar**. Se trata de una conclusión evidente, porque de momento somos los humanos los que utilizamos la nueva tecnología para explorar resultados concretos para una **tarea determinada que debemos definir**.

Revision #11

Created 2026-03-10 20:01:00 CET by Miguel Serrano Larraz

Updated 2026-03-16 18:02:22 CET by Miguel Serrano Larraz