



**Manuel Torres**  
AI Manager - Data Intelligence

**Manuel Ángel García Sánchez**  
Manager Inteligencia de Negocio



# Parole, parole, parole... o cómo se consigue que las máquinas nos entiendan

En una sociedad de la información como la nuestra, tener **sistemas** informáticos que **puedan interactuar** con nosotros, **aprender** de la información y ayudarnos a **entender** la gran cantidad de información disponible es clave para el éxito. El **lenguaje es nuestra principal forma de comunicación**. Lo usamos para interactuar entre nosotros, para aprender, para expresar nuestros deseos, sentimientos, órdenes e ideas.

El **Procesamiento del Lenguaje Natural** (NLP por sus siglas en inglés) es el área de la **Inteligencia Artificial** que trata la interacción entre los sistemas informáticos y el lenguaje humano. **Asistentes inteligentes, chatbots, traductores automáticos...** hay muchas nuevas tecnologías construidas a partir de la Inteligencia Artificial que utilizan el Procesamiento del Lenguaje Natural como su base.

La **información no estructurada** representa en torno al **90% del total de datos** de una empresa, según el informe BI Trend Monitor 2016 del Business Application Research Center (BARC). Con esta prevalencia, uno de los retos principales de las organizaciones **es incluir la gestión y el tratamiento de estos datos entre sus prioridades más acuciantes**.

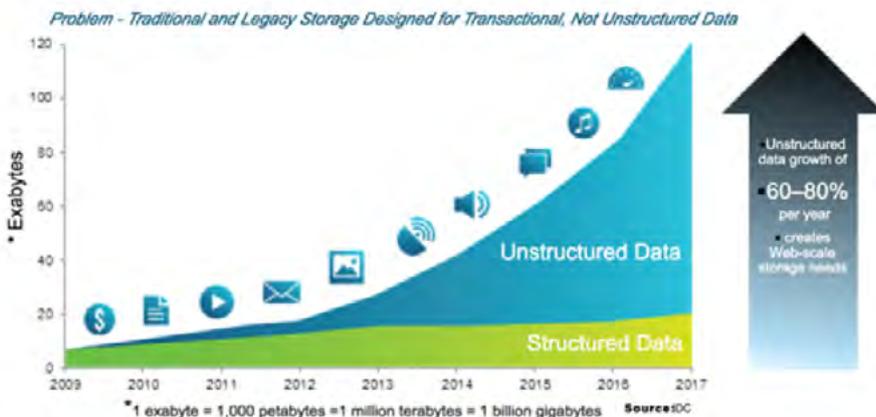
El procesamiento del lenguaje natural se especializa en el **procesamiento, transcripción, comprensión y generación del lenguaje humano**. Hay diversos campos en el NLP que abordan diferentes áreas del procesamiento del lenguaje humano: el habla, la comprensión del lenguaje y la generación del lenguaje. Estos campos utilizan diferentes técnicas y diferentes enfoques. En conjunto, se utilizan para dar a sistemas automáticos la capacidad de comunicarse de manera natural con interlocutores humanos, bien para entender órdenes, para procesar el contenido de textos o para generar mensajes que sean más amigables y comprensibles.

Uno de los enfoques es el uso de técnicas de **Text Mining** que ayudan a la **compresión, clasificación y análisis de textos**.

Extraer todo el potencial de esas fuentes de datos puede ser una tarea complicada debido a múltiples factores:

- Grandes volúmenes de datos y formatos inconsistentes.
- Múltiples fuentes e idiomas.
- Errores de ortografía, jerga y abreviaturas
- Alta subjetividad a la hora de la interpretación y al contexto.

## Data Growth



El término **analítica de textos** (o minería de textos) describe un conjunto de técnicas lingüísticas, estadísticas y de aprendizaje automático que modelan y estructuran el contenido de fuentes de datos que permiten realizar análisis, búsquedas o investigación.

Existen tres técnicas principales en el Text Mining:

- **Extracción de entidades:** mediante esta técnica se identifican los diferentes elementos y entidades lógicas. El resultado es el listado de todos los términos (ponderados) que aparecen en el texto a analizar.

- **Extracción de información:** trata de identificar las relaciones básicas de los términos previamente extraídos. Como puede ser las funciones que representan cada uno de ellos dentro del conjunto global del texto.
- **Análisis relacional:** permite determinar las relaciones entre los diferentes términos con múltiples conexiones. Este tipo de técnicas ayudan a entender el contenido de un texto y a su categorización.

El uso de las técnicas de Text Mining permite solucionar múltiples problemas habituales en las empresas:

- **Análisis de documentos legales:** identificación de los diferentes elementos de un contrato, identificación de fechas de vencimiento, detección de cláusulas abusivas, etc.
- **Categorización de buzones de reclamaciones e incidencias:** identificando el contenido de cada uno de los correos se puede redirigir al departamento encargado de su tratamiento ayudando, no sólo a reducir costes, sino también eliminando posibles subjetividades a la hora del tratamiento de estos por los diferentes componentes humanos de los equipos encargados.
- **Análisis de resultados de campañas de marketing** mediante análisis de sentimientos.

Otro de los enfoques es el uso de **interfaces conversacionales de usuario:**

**Chatbots y Asistentes Inteligentes.** Más de 4 mil millones de usuarios globales son usuarios de plataformas de mensajería. Seis de las 10 aplicaciones más utilizadas, son aplicaciones de mensajería. Todos las usamos diariamente y nos hemos acostumbrado a trabajar con ellas de manera natural. Los chatbots permiten extender esta forma de comunicación y utilizarla como interfaz de usuario para los sistemas informáticos. Estos interfaces permiten establecer un diálogo (en lenguaje natural) entre usuarios y sistemas, permitiendo interfaces más flexibles y ricas.

Se consideran como una extensión de la experiencia de usuario, permitiendo el acceso a usuarios o clientes a través de más canales. La comunicación puede ser mediante voz (convertida a texto usando sistemas de Machine Learning) o directamente mediante texto. Una vez en formato de texto, los Chatbots realizan diferentes tareas de procesamiento de lenguaje:



**Embeddings.** Proporcionan una manera de utilizar una representación numérica eficiente y densa en la que las palabras similares tienen una codificación similar (están cerca unas de otras) desde el punto de vista de los modelos de NLP. La representación es generada por modelos de Deep Learning que analizan un gran corpus de texto y producen una representación vectorial de cada palabra que conserva la relación de esa palabra con el resto del vocabulario.

**Reconocimiento de entidades:** Técnica habitual en los procesos de Text Mining (como mencionamos anteriormente) que permite la extracción de una secuencia de elementos dentro del texto (personas, organizaciones, ubicaciones, marcas comerciales, etc.).

**Detección de intenciones:** La detección de intenciones es una técnica de procesamiento de lenguaje que permite determinar cuál es el propósito de cada frase y agrupar las frases que tienen el mismo propósito para extraer la intención del hablante. En los chatbot, las intenciones se hacen corresponder con comandos que se deben ejecutar y que corresponden a las funcionalidades que se desean proporcionar mediante su uso.

**Análisis de sentimientos:** El análisis de sentimiento o minería de opinión (opinion mining) es el proceso de determinar el tono emocional que hay detrás de una serie de palabras. Se utiliza para intentar entender las actitudes, opiniones y emociones expresadas.

**Procesamiento de acciones:** Una vez identificados los términos, intenciones y el tono de la petición, el chatbot se encarga de identificar las acciones a llevar a cabo y por último, la generación de la respuesta al interlocutor.

En el desarrollo de un chatbot es fundamental pensar en la regla del 80/20: el **80% del tiempo** se debería dedicar a la **creación y mantenimiento de contenidos** y el 20% se emplearía en tareas de inteligencia artificial (IA) y machine learning (ML). Lo importante siempre debe ser qué funcionalidad exponemos, no el desarrollo de IA. Es por ello que el rol del **“Product Owner”** sea clave para dotar los contenidos adecuados al chatbot.

En conclusión, destacar que las aplicaciones del procesamiento del lenguaje natural permiten integrar todo el poder comunicativo del habla humana con las capacidades de procesamiento de los sistemas informáticos.

