



NOMBRE MIGUEL ANGEL GUEVARA SANTILLAN

NOMBRE DEL MAESTRO: ING.JOSE JULIO GONZALEZ ALVAREZ

MATERIA: INTELIGENCIA ARTIFICIAL

GRUPO: 701 MIXTO

# PLN o Procesamiento de Lenguaje Natural

El **Procesamiento del Lenguaje Natural** es una rama de conocimiento de la Inteligencia Artificial que, esencialmente, pretende conseguir que una máquina comprenda lo que expresa una persona mediante el uso de una lengua natural (inglés, español, chino...). Las lenguas naturales pueden expresarse de forma oral (mediante la voz), escrita (un texto) o por signos. Obviamente, la expresión escrita está mucho más documentada y es más fácil de conseguir y tratar que la oral o el lenguaje de signos. Por lo tanto, el Procesamiento del Lenguaje Natural está mucho más avanzado en el tratamiento de textos escritos.

En teoría, cualquier lengua humana puede ser interpretada y tratada por un ordenador, aunque esto en la práctica no se hace, dadas las limitaciones económicas o las escasas aplicaciones que tendría interpretar las más de 3.000 lenguas catalogadas existentes en el mundo.

## Modelos para el Procesamiento de Lenguaje Natural

Que una máquina se comunique con una lengua natural implica tratar de manera computacional esa lengua. ¿Cómo se transforma un lenguaje natural en lenguaje computacional? Esto requiere un proceso de modelización matemática, para que un ordenador, que solo entiende bytes y dígitos, comprenda el lenguaje humano.

Por un lado, los informáticos codifican los programas empleando lenguajes de programación, y por otro, los lingüistas computacionales se encargan de preparar el **modelo lingüístico** para que los ingenieros informáticos lo implementen en un código eficiente y funcional.

Los modelos lingüísticos utilizados en PLN son esencialmente dos: los modelos lógicos, basados en gramáticas, y los modelos probabilísticos, basados en datos.

## Modelos lógicos

En los modelos lógicos, los lingüistas escriben reglas de reconocimiento de patrones estructurales, utilizando un formalismo gramatical concreto. Los patrones estructurales son definidos por estas reglas y por la información almacenada en diccionarios computacionales.

## **Modelos probabilísticos**

En los modelos probabilísticos, el proceso es a la inversa: los lingüistas recogen un corpus de ejemplos y datos y, a partir de estas colecciones, se calcula la frecuencia con la que las unidades lingüísticas aparecen en un contexto determinado (por ejemplo, frecuencia en la que aparece la palabra "tarjeta" en la colección de ejemplos y datos). Con este cálculo, podemos predecir cuál será la siguiente unidad en un contexto dado (siguiendo con el ejemplo anterior, se podría calcular la frecuencia con la que la palabra "tarjeta" aparece unida a "sanitaria"). Esto es el paradigma de "**aprendizaje automático**": los algoritmos infieren las posibles respuestas a partir de los datos observados anteriormente en el corpus.

## **Componentes del Procesamiento del Lenguaje Natural**

Los principales componentes del Procesamiento del Lenguaje Natural son el análisis morfológico o léxico, el Análisis sintáctico, el Análisis semántico y el Análisis pragmático. Cada tipo de análisis (o la combinación de varios) se utilizará según sea el fin de su aplicación. Por ejemplo, para un sistema conversacional necesitaríamos un análisis pragmático, que observará con precisión todos los detalles del contexto.

Puedes leer el artículo completo sobre Procesamiento de Lenguaje Natural, definición, modelos y componentes en el [blog del IIC](#).