

El Procesamiento del Lenguaje Natural es una rama de conocimiento de la Inteligencia Artificial que, básicamente, pretende conseguir que una máquina comprenda lo que una persona expresa mediante el uso de su lengua natural de expresión (idioma inglés, español, chino, etc.)

También puede definirse al procesamiento de lenguajes naturales —abreviado PLN, o NLP del idioma inglés *Natural Language Processing*— como un campo de las ciencias de la computación, inteligencia artificial y lingüística que estudia interacciones entre las computadoras y el lenguaje humano.

Las lenguas naturales pueden expresarse en tres formas: oral (mediante la voz), escrita (un texto) o por signos. De más está decir que la expresión escrita se encuentra mucho más documentada y es más fácil de conseguir y analizar que la oral o el lenguaje de signos. Por ello, el Procesamiento del Lenguaje Natural está mucho más avanzado en el tratamiento de textos escritos.

En teoría, cualquier lengua humana debería estar en condiciones de ser interpretada y tratada por un ordenador. No obstante, en la práctica no ocurre. ¿A qué se debe? A las limitaciones económicas o los escasos usos que tendría la tarea de interpretar las más de 3.000 lenguas catalogadas que existen en el mundo.

El PLN no se ocupa de la comunicación por medio de lenguajes naturales de una forma abstracta. Busca diseñar mecanismos para comunicarse que sean eficaces computacionalmente, esto es, que se puedan realizar por medio de programas que ejecuten o simulen la comunicación.

Los modelos aplicados se enfocan no solo a la comprensión del lenguaje de por sí, sino a aspectos generales cognitivos de los humanos y a la organización de su memoria. El lenguaje natural es solo un medio para estudiar estos fenómenos.

Hasta la década del 80, la mayoría de los sistemas de PLN estaban basados en un complejo conjunto de reglas diseñadas manualmente. A partir de finales de 1980, se produce una revolución en ese campo, a partir de la introducción de algoritmos de aprendizaje automático para el procesamiento del lenguaje.

Idiomas como el inglés, español, alemán, francés, portugués, chino, árabe y japonés (el orden de exposición es aleatorio) son los que acumulan más aplicaciones. Google Translate es la herramienta que más lenguas interpreta. Supera el centenar, pero se estima que hay entre 5.000 y 7.000 lenguas en el mundo

El hecho de que una máquina se comunique con una lengua natural implica que la misma sea tratada de manera computacional. ¿Cómo se transforma un lenguaje natural en lenguaje computacional? Requiere de un proceso de modelización matemática para que un ordenador, que solo entiende de bytes y dígitos, comprenda el lenguaje humano.