

## Tema 2. Utilización de modelos de Inteligencia Artificial

Id	Ambiente	Temperatura	Humedad	Viento	Clase
E1	Soleado	Alta	Alta	Falso	No
E2	Soleado	Alta	Alta	Verdadero	No
E3	Nublado	Alta	Alta	Falso	Sí
E4	Lluvioso	Media	Alta	Falso	Sí
E5	Lluvioso	Baja	Normal	Falso	Sí
E6	Lluvioso	Baja	Normal	Verdadero	No
E7	Nublado	Baja	Normal	Verdadero	Sí
E8	Soleado	Media	Alta	Falso	No
E9	Soleado	Baja	Normal	Falso	Sí
E10	Lluvioso	Media	Normal	Falso	Sí
E11	Soleado	Media	Normal	Verdadero	Sí
E12	Nublado	Media	Alta	Verdadero	Sí
E13	Nublado	Alta	Normal	Falso	Sí
E14	Lluvioso	Media	Alta	Verdadero	No

Tabla 1. Instancias con sus valores de atributos y clases del problema 'Jugar al aire libre'. Fuente: elaboración propia.

### Factores de certeza

El uso de factores de certeza es una alternativa al razonamiento bayesiano. El primer sistema que utilizó estos factores fue el sistema experto MYCIN (Shortliffe & Buchanan, 1975), escrito en LISP, y que permitía diagnosticar infecciones de sangre y meningitis. Dado que no había datos estadísticos fiables sobre el problema, los creadores de MYCIN utilizaron un factor de certeza para indicar la confianza o certeza del experto humano.

El **factor de certeza** puede tener un valor entre -1 y +1. -1 indica falsedad total, mientras que el valor +1 indica certeza total.

Un valor positivo indica cierto grado de creencia en la expresión, mientras que un valor negativo indica cierto grado de incredulidad respecto a la veracidad de la expresión.

## Tema 2. Utilización de modelos de Inteligencia Artificial

Si se emplean factores de certeza, las reglas en la base de conocimiento irán acompañadas de este factor, que indica el grado de certeza de que se dé el hecho expresado en el consecuente o hipótesis, dado el hecho expresado en el antecedente o evidencia.

Cuando se trabaja con factores de certeza se pueden encontrar reglas como la que sigue:

SI 'vivienda en propiedad'

ENTONCES 'sistema económico es solvente' {cf 0.6}

'sistema económico es insolvente' {cf 0.2}

Entre paréntesis, junto a cada consecuente, se representa el factor de certeza precedido de las siglas *cf* (*certainty factor*). Ese valor, por tanto, representa la creencia en que se dé la hipótesis, dada la evidencia.

En el ejemplo, dada la evidencia 'vivienda en propiedad', se tiene una mayor certeza (0.6) de que el sistema económico sea solvente a que sea insolvente (0.2). No es imprescindible que ambos factores de certeza sumen 1, como se refleja en el ejemplo. Puede existir un factor 0.2 que se podría asignar a otro hecho como un término medio entre solvente e insolvente, u otro valor no observado.

### ¿Cómo se propagan los factores de certeza en el encadenamiento de reglas?

Puede suceder que el propio antecedente de la regla sea incierto y tenga asignado un factor de certeza y este factor se ha de propagar a lo largo de la cadena de reglas.

## Tema 2. Utilización de modelos de Inteligencia Artificial

El factor de certeza para una regla con un único antecedente se puede calcular simplemente multiplicando el factor de certeza del antecedente (la evidencia  $E$ ) por el factor de certeza de la regla cuyo consecuente es  $H$  mediante la siguiente fórmula:

$$cf(H, E) = cf(E) \cdot cf(r)$$

Si en el anterior ejemplo, el factor de certeza de que la 'vivienda en propiedad' es 0.6, entonces se tienen los siguientes factores de certeza para cada una de las hipótesis:

$$cf(H_1, E) = 0.6 \cdot 0.6 = 0.36$$

$$cf(H_2, E) = 0.2 \cdot 0.6 = 0.12$$

Siendo  $H_1$  la hipótesis 'sistema económico es solvente' y  $H_2$ , la hipótesis 'sistema económico es insolvente'.

En el caso de que existan múltiples antecedentes, el factor de certeza se propaga de diferente manera según las evidencias o antecedentes sean una conjunción (AND) o disyunción (OR).

En el caso de que sea una regla conjuntiva del tipo:

SI     $E_1$

AND  $E_2$

AND  $E_3$

...

AND  $E_n$

ENTONCES

$H_1$     {cf}

## Tema 2. Utilización de modelos de Inteligencia Artificial

El factor de certeza del consecuente, la hipótesis, se calcula de la siguiente manera:

$$cf(H, E_1 \cap E_2 \cap E_3 \cap \dots \cap E_n) = \min[cf(E_1), cf(E_2), \dots, cf(E_n)] \cdot cf(H)$$

Para el siguiente ejemplo:

SI 'vivienda en propiedad'

AND 'ingresos altos'

ENTONCES 'sistema económico es solvente' {cf 0.6}

Si el factor de certeza del hecho 'vivienda en propiedad' es 0.7 y el factor de certeza del hecho 'ingresos altos' es 0.4, entonces el factor de certeza del consecuente se calcula de la siguiente manera:

$$cf(H, E_1 \cap E_2) = \min[0.7, 0.4] \cdot 0.6 = 0.24$$

En el caso de que se tenga una regla disyuntiva del tipo:

SI E1

OR E2

OR E3

...

OR En

ENTONCES

H1 {cf}

El factor de certeza del consecuente, la hipótesis, se calcula de la siguiente manera:

$$cf(H, E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n) = \max[cf(E_1), cf(E_2), \dots, cf(E_n)] \cdot cf(H)$$

## Tema 2. Utilización de modelos de Inteligencia Artificial

Para el siguiente ejemplo:

SI 'vivienda en propiedad'

OR 'ingresos altos'

ENTONCES 'sistema económico es solvente' {cf 0.6}

Si el factor de certeza del hecho 'vivienda en propiedad' es 0.7 y el factor de certeza del hecho 'ingresos altos' es 0.4, entonces el factor de certeza del consecuente se calcula de la siguiente manera:

$$cf(H, E_1 \cup E_2) = \max[0.7, 0.4] \cdot 0.6 = 0.42$$

Los factores de certeza y su propagación a lo largo de las reglas encadenadas, aunque no tienen gran base matemática como el razonamiento bayesiano, son muy prácticos en algunos problemas produciendo buenos resultados con menos coste computacional. Además, los métodos estadísticos requieren el conocimiento de gran cantidad de datos.

### Lógica difusa o 'Fuzzy logic'

La **lógica difusa** es una familia de teorías y técnicas basadas en el concepto de conjuntos difusos, también denominados *conjuntos borrosos*.

Mientras que la teoría de conjuntos tradicional define 'ser miembro de un conjunto' como un predicado booleano, la teoría de conjuntos difusos permite representar el "ser miembro de un conjunto" como una distribución de posibilidades.

El padre de la lógica difusa es Zadeh (1965), el cual afirma que "a medida que la complejidad de un problema aumenta, disminuye la posibilidad de analizarlo en términos precisos". Es decir, **a medida que el problema es más complejo, es más difícil construir métodos cuantitativos.**

La lógica difusa es una extensión de la lógica tradicional, la cual resulta natural y útil,

## Tema 2. Utilización de modelos de Inteligencia Artificial

porque las personas normalmente razonamos bien con términos imprecisos. Al perder el enfoque booleano, se cubren más aspectos con una regla y se razona de forma sencilla pero más potente.

La lógica difusa ha sido aplicada en áreas tan diversas como control, medicina, biología, ecología, economía o política. En la ingeniería de control se aplicó por primera vez en 1974 (Mamdani y su grupo del Queen Mary College en el Reino Unido lo aplicaron para el control de una máquina de vapor).

Los sistemas basados en lógica difusa pueden **controlar más adecuadamente procesos que estén gobernados por reglas intuitivas** que difícilmente pueden expresarse matemáticamente.

La gran potencia de esta metodología programable se debe a la posibilidad de expresar operaciones y controlar las reglas del sistema mediante palabras de uso cotidiano. Por ejemplo, en un sistema de control de un ascensor, podría programarse:

SI está cerca de un piso AND hay orden de parar

ENTONCES disminuir la velocidad

En este caso, una entrada al sistema de control sería la posición del ascensor y, como 'cerca' es un conjunto difuso (concepto que se explicará a continuación), el valor de verdad de la premisa y, por tanto, el de la velocidad, varían de acuerdo con dicha posición.

La forma de expresar las reglas de operación mediante palabras permite controlar procesos sencillos con una decena de reglas y procesos complejos con 30 o 40, reduciendo considerablemente la cantidad de código de programación y, por tanto, el tiempo de diseño, el tiempo de desarrollo de un prototipo, la carga computacional y de memoria, etc.

## Tema 2. Utilización de modelos de Inteligencia Artificial

Otra ventaja del control difuso es la **fácil modificación del funcionamiento del sistema**, que puede llevarse a cabo cambiando algunas premisas y operaciones o añadiendo reglas (el criterio de comportamiento del sistema va implícito en las reglas). Sin embargo, en un sistema convencional, un pequeño cambio puede requerir la derivación completa de nuevas ecuaciones. El control difuso no necesita de una etapa previa de obtención del modelo matemático del proceso.

### Conjuntos difusos

En lógica tradicional la pertenencia a un conjunto es binaria (pertenece o no pertenece). La lógica difusa extiende el concepto de pertenencia asignando una probabilidad de pertenencia (ver figura).

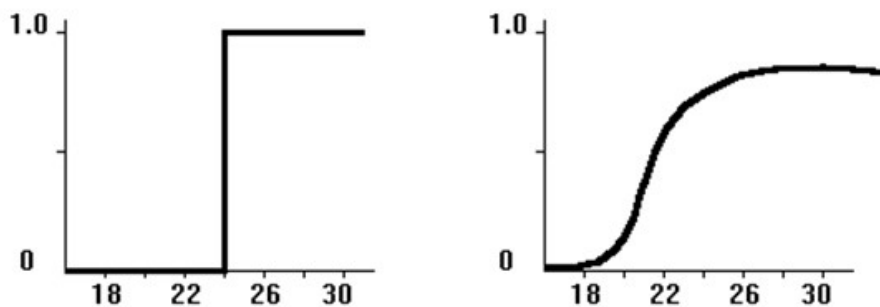


Figura 12. Pertenencia a un conjunto en lógica tradicional (gráfico de la izquierda).

Se muestran a continuación ejemplos para ilustrar las diferentes lógicas. Primero se considera el conjunto X de los números reales entre 0 y 10, que se llamará **conjunto universal**. Después, se define un subconjunto A de X de todos los números reales que están en el rango entre 5 y 8.

$$A = [5, 8]$$

## Tema 2. Utilización de modelos de Inteligencia Artificial

La función característica del conjunto  $A$  asigna un número, 1 o 0 para cada elemento en  $X$ , y el valor depende de si el elemento pertenece al subconjunto de  $A$  o no. Este resultado se muestra en la figura:

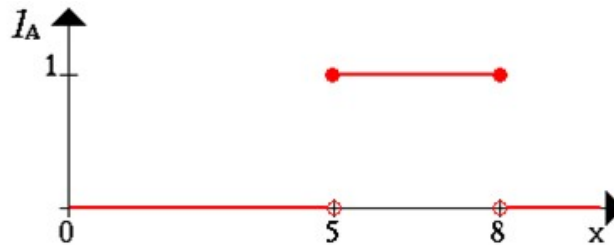


Figura 13. Ejemplo de pertenencia a un conjunto en lógica tradicional.

Los elementos del conjunto  $A$  tienen asignados el valor de 1 porque pertenecen a ese conjunto y cada elemento que tiene asignado el número 0 es un elemento que no está en el conjunto  $A$ .

Este concepto es suficiente para aplicaciones en muchas áreas, pero se necesita buscar solución a situaciones donde se requiere flexibilidad. Para ilustrar cómo modelar esta flexibilidad, se considera el siguiente ejemplo que consiste en describir a la gente de peso saludable.

Formalmente se puede denotar así:

$A = \{\text{Conjunto de personas de peso normal}\}$

Entonces, en general, se toma como límite inferior las personas con Índice de Masa Corporal (IMC) igual a 18.5 mientras que como límite superior se utiliza un IMC igual a 25. Entonces, el conjunto  $A$  se define como un intervalo abrupto (Ver figura):

$A = [18.5, 25]$

## Tema 2. Utilización de modelos de Inteligencia Artificial

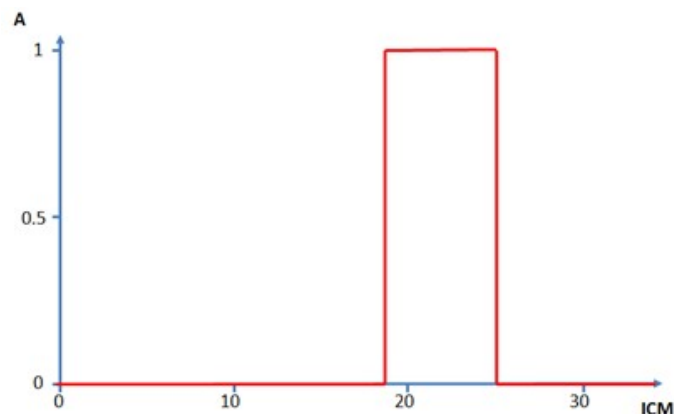


Figura 14. Representación del conjunto abrupto 'personas de peso normal'.

Surge ante esta representación la pregunta siguiente: ¿Una persona con IMC igual a 25 tiene un peso normal mientras que una persona con IMC igual a 25,1 tiene sobrepeso? Si se utilizan conjuntos abruptos, un sistema sí asignaría de esa manera los pesos, mientras que si se definen conjuntos difusos se puede definir un grado de peso y un grado de sobrepeso para las personas cuyo IMC está cerca del límite establecido.

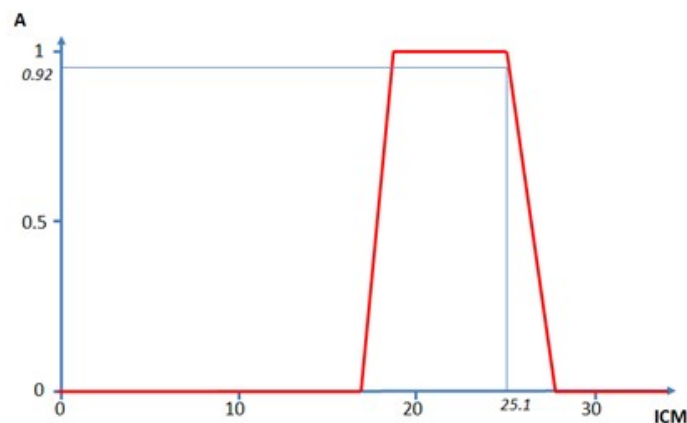


Figura 15. Representación del conjunto difuso 'personas de peso normal'.

Tal y como se muestra, hay valores de IMC a los que se les asigna un valor de la función entre 0 y 1, que indica el grado de pertenencia de cada elemento al conjunto A. Si está en 1 significa que el elemento pertenece al conjunto A totalmente, mientras

## Tema 2. Utilización de modelos de Inteligencia Artificial

que si está en 0 significa que el elemento no pertenece al conjunto A. Para concretar en este ejemplo, si la persona tiene un IMC igual a 25.1 se le considera una pertenencia del 92% al conjunto de personas de peso normal.

La representación del conjunto difuso viene establecida por la función de pertenencia.

**Función de pertenencia (m):** Función que define el conjunto difuso A en el universo U. Para un valor x del universo U, la función de pertenencia  $m_A(x)$  indica el grado de pertenencia de x al conjunto A, pudiendo ser este grado un valor entre 0 y 1 inclusive. Esto se denota como:

$$\mu:U \rightarrow [0,1]$$

A **m** también se le conoce como el **valor de verdad** porque representa el grado en que una proposición es verdadera.

### Variables lingüísticas

Una variable lingüística es una variable cuyos valores son términos lingüísticos y se define como un quinteto (X, T(X), U, G, M), donde:

- ▶ **X:** nombre de la variable.
- ▶ **T(x):** conjunto de valores lingüísticos (atributos, adjetivos) de x.
- ▶ **U:** universo de discurso (rango de posibles valores de la variable x).
- ▶ **G:** regla sintáctica para generar los valores lingüísticos de x.
- ▶ **M:** regla semántica para asociar a cada término lingüístico su significado, que es un conjunto difuso en U. Cada conjunto difuso, por tanto, representa un valor lingüístico de la correspondiente variable lingüística.

# Tema 2. Utilización de modelos de Inteligencia Artificial

Por ejemplo (ver figura):

- ▶  $X = T = \text{Temperatura}$
- ▶  $T(x) = \{\text{baja, muy baja, moderadamente alta, ...}\}$
- ▶  $U = [100^\circ \text{C}, 500^\circ \text{C}]$
- ▶ G: el término lingüístico “baja” se utiliza para los valores de T por debajo de  $10^\circ \text{C}$
- ▶ M: el conjunto difuso representando el valor “moderado” se define para una temperatura alrededor de  $250^\circ \text{C}$  con una función de pertenencia  $\mu_{\text{moderado}}$ .

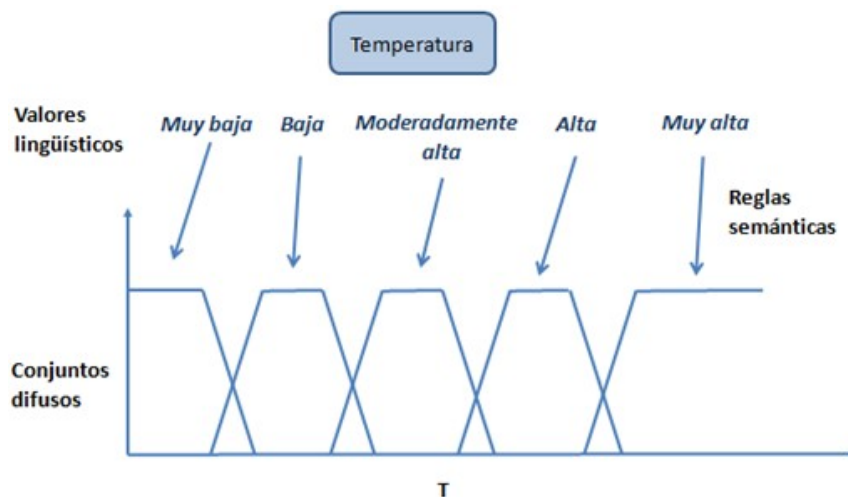


Figura 16. Variable lingüística temperatura.

A los conjuntos difusos se les puede aplicar unos **modificadores lingüísticos** (*hedge*), que son funciones matemáticas que modelan un adverbio lingüístico, modificando la forma de los conjuntos difusos. Se utilizan para cuantificar las etiquetas de una variable lingüística: muy, algo, bastante, poco, etc.

## Tema 2. Utilización de modelos de Inteligencia Artificial

Existen definiciones de modificadores lingüísticos comúnmente utilizadas como, por ejemplo, las mostradas en la siguiente figura:

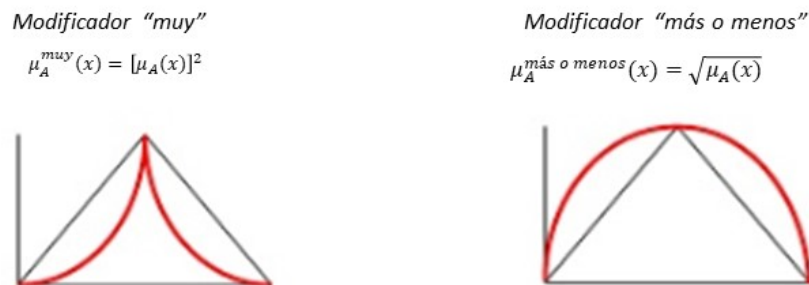


Figura 17. Modificadores lingüísticos "muy" y "más o menos".

Como se puede observar en la figura, el modificador "muy" reduce el grado de pertenencia al conjunto modificado. Así, si la función está representando la altura y a un hombre le corresponde un grado de pertenencia de 0.9 al conjunto de "personas altas", le correspondería un grado de pertenencia de  $0.9^2=0.81$  al conjunto de "personas muy altas".

Por el contrario, como se puede observar en la figura anterior, el modificador "más o menos" realiza una operación de expansión, relajando la condición de pertenencia al conjunto y, por tanto, se aumenta el grado de pertenencia de los elementos al conjunto. Si este operador es aplicado al ejemplo de la altura previo, el hombre tiene una pertenencia al conjunto de "personas más o menos altas" de 0.94.

### Reglas difusas

Las variables lingüísticas se utilizan en reglas difusas. Una regla difusa consiste en una expresión:

SI  $x$  es  $a$

ENTONCES  $y$  es  $b$

## Tema 2. Utilización de modelos de Inteligencia Artificial

Donde  $x$  e  $y$  son variables lingüísticas y  $a$  y  $b$  son valores lingüísticos que vienen determinados por conjuntos difusos definidos en el universo de discurso de las variables  $x$  e  $y$ , respectivamente.

SI            temperatura            es            alta

ENTONCES    riesgo de avería            es            moderado

El rango de la variable temperatura (el universo de discurso) está establecido entre 100 y 500° C. En la regla se especifica el valor lingüístico “alta” correspondiente a uno de los conjuntos difusos definidos mediante funciones de pertenencia a lo largo de ese universo de discurso.

Otros conjuntos difusos corresponderán a los valores “baja”, “moderadamente alta” o “muy alta”, por ejemplo. Igualmente, en el consecuente se encuentra una variable lingüística con un valor lingüístico asociado a un determinado conjunto difuso.

A diferencia de las reglas clásicas o no difusas (si no se emplean mecanismos tales como factores de certeza), en las que, si el antecedente se cumple, entonces el consecuente se cumple y se considera totalmente cierto, en las reglas difusas el antecedente se cumple parcialmente y este grado de veracidad se propaga al consecuente y, por tanto, el consecuente también será parcialmente veraz en un grado análogo.

Tanto antecedentes como consecuentes de las reglas difusas pueden tener múltiples condiciones. En el caso de los antecedentes se han de aplicar operadores para generar un valor que afectará por igual a todos los consecuentes. Estos operadores son denominados **métodos de implicación**.

# Tema 2. Utilización de modelos de Inteligencia Artificial

## 2.5. Referencias bibliográficas

Bornet, P., Barkin, I. y Wirtz, P. (2020). *Intelligent Automation - Learn How to Harness Artificial Intelligence to Boost Business & Make Our World More Human*.

Departamento de CCIA de la Universidad de Sevilla. (s.f.). *Sistemas Basados en reglas*.

[https://www.cs.us.es/~fsancho/Blog/posts/Sistemas\\_Basados\\_en\\_reglas.md.html](https://www.cs.us.es/~fsancho/Blog/posts/Sistemas_Basados_en_reglas.md.html)

Oracle Netsuite. (2020, diciembre 2). *32 Business Automation Statistics for 2021*.

<https://www.netsuite.com/portal/resource/articles/business-strategy/business-automation-statistics.shtml>

Wang, L., Lin, Z. Q. y Wong, A. (2020). COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports*, 10(19549).

# Tema 3. Procesamiento del Lenguaje Natural

## 3.1. Introducción y objetivos

Este tema comienza introduciendo de manera general el campo de la inteligencia artificial denominado procesamiento del lenguaje natural (PLN). La idea de que las máquinas sean capaces de poseer habilidades de comunicación y lenguaje es tan antigua como la propia aparición de los primeros ordenadores.

En los siguientes apartados se describe la historia del procesamiento del lenguaje natural y se puede observar cómo ha evolucionado dicho campo y los diferentes enfoques que ha ido tomando a lo largo de la historia. Por último, analizamos cómo el conocimiento del lenguaje es imprescindible para el correcto funcionamiento de los sistemas de procesamiento de lenguaje natural, describiendo los diferentes tipos de conocimiento del lenguaje.

# Tema 3. Procesamiento del Lenguaje Natural

## 3.2. Procesamiento del Lenguaje Natural – PLN

### ¿Qué contempla el PLN?

El procesamiento del lenguaje y del habla ha sido tratado históricamente de forma muy diferente en la informática, la ingeniería, la lingüística, la psicología o la ciencia cognitiva. Hoy en día se concibe el procesamiento del lenguaje natural como área que abarca varios campos diferentes y diversos pero superpuestos. Por ello, el procesamiento del lenguaje natural es un campo interdisciplinario que une a informáticos, ingenieros electrónicos y de telecomunicaciones con lingüistas, sociólogos y psicólogos.

El reconocimiento de la voz, que incluye tareas del procesamiento de la señal, se ha tratado tradicionalmente en la ingeniería electrónica y de telecomunicaciones y es una parte del PLN. El **análisis sintáctico** y la **interpretación semántica** de las palabras y frases son áreas tradicionales del procesamiento del lenguaje natural que se estudian en el campo de la informática. La **morfología**, la **fonología** y la **pragmática** son tareas de investigación en la lingüística computacional. **Psicolingüistas y sociolingüistas estudian respectivamente los mecanismos cognitivos para la adquisición del lenguaje y cómo la sociedad influye en el uso de la lengua.**

El ancho espectro que abarca el campo del procesamiento del lenguaje natural hace que se conozca con diferentes nombres debido a las diferentes vertientes involucradas. Algunos de estos nombres, que provienen de estas diferentes facetas, serían procesamiento del lenguaje y del habla, tecnología del lenguaje, procesamiento del lenguaje natural, lingüística computacional o reconocimiento y síntesis del habla.

# Tema 3. Procesamiento del Lenguaje Natural

En la inteligencia artificial, el procesamiento del lenguaje natural es un campo que tiene como objetivo que las máquinas sean capaces de realizar tareas que involucren el lenguaje humano.

Algunas de las tareas que debe realizar una máquina para ser capaz de procesar el lenguaje natural incluyen funcionalidades tales como la de habilitar a la máquina de habilidades para comunicarse con personas, la de mejorar la comunicación entre humanos o, simplemente, la de procesar un texto o el habla.

A pesar de ser un campo aún experimental, el procesamiento del lenguaje natural es una de las tecnologías que se esconde tras la inteligencia artificial con mayor proyección.

## Historia del procesamiento del lenguaje natural

La historia del procesamiento del lenguaje natural se puede dividir en diferentes etapas. Desde la aparición de las bases o paradigmas fundacionales en la década de 1940 hasta la explotación de los paradigmas más modernos para desarrollar hoy en día aplicaciones más inteligentes.

1940-1950	Paradigmas fundacionales
1957-1970	Paradigma simbólico y estocástico
1970-1983	Cuatro paradigmas de investigación
1983-1993	Revivir del empirismo y los modelos de estados finitos
1994-1999	Unión de las diferentes vertientes
2000	Auge del aprendizaje automático

Tabla 1. Cronología de las diferentes etapas de la historia del PLN. Fuente: elaboración propia.

# Tema 3. Procesamiento del Lenguaje Natural

## Paradigmas fundacionales: década de 1940 y 1950

El principio del PLN data del período justo después de la II Guerra Mundial, cuando se dio el origen del ordenador. En este período, desde la década de 1940 hasta el final de la década de 1950, se trabajó intensamente en dos paradigmas fundacionales:

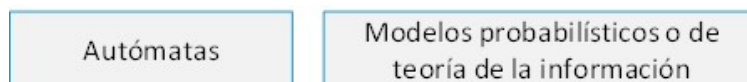


Figura 1. Paradigmas fundacionales. Fuente: elaboración propia.

### Autómatas

El autómata surgió en la década de 1950 a partir del famosísimo estudio publicado por Turing (1936), *Los números computables, con una aplicación a un problema de decisión*, y que se considera como la base de la informática moderna. El trabajo de Turing condujo primero a un modelo simplificado de la neurona como elemento de computación que podría describirse en términos de lógica proposicional (McCulloch y Pitts, 1943) y luego a la definición de los **autómatas finitos** y las **expresiones regulares** (Kleene, 1951).

Shannon aplicó las cadenas de Markov, un modelo de proceso estocástico discreto en el que la probabilidad de ocurrencia de un evento depende solo del evento inmediatamente anterior, a los autómatas para el lenguaje (Shannon, 1948). Aprovechando las ideas del trabajo de Shannon, Chomsky fue el primero en considerar las máquinas de estados finitos como una forma de caracterizar una gramática y definió el lenguaje de estados finitos como un lenguaje generado por una gramática de estados finitos (Chomsky, 1956).

Estos primeros modelos llevaron a la aparición de la **teoría del lenguaje formal**, que utilizó el álgebra y la teoría de conjuntos para definir los lenguajes formales como

## Tema 3. Procesamiento del Lenguaje Natural

secuencias de símbolos. Esta teoría incluye las gramáticas libres de contexto, un concepto que Chomsky definió en 1956 para las lenguas naturales, pero que también fue descubierto de forma independiente por Backus y Naur en su descripción del lenguaje de programación ALGOL (Backus, 1959) (Naur et al., 1960).

### Teoría de la información

El segundo paradigma fundamental de este período fue el desarrollo de algoritmos probabilísticos para el procesamiento del lenguaje y del habla. Esta idea proviene de la otra gran contribución de Shannon, el teorema de codificación de canal en la **teoría de la información**, y que muestra que es posible la transmisión y decodificación del lenguaje a través de un canal de comunicación ruidoso.

Shannon tomó prestado el concepto de **entropía** de la termodinámica como una forma de medir la capacidad de información de un canal o el contenido de información de un idioma, y realizó la primera medida de la entropía del inglés utilizando técnicas probabilísticas.

También durante este período inicial se desarrolló el espectrógrafo de sonido y se realizó investigación fundamental en la fonética instrumental, lo que sentó las bases para el **reconocimiento de la voz**. La primera máquina capaz de realizar el reconocimiento de la voz apareció a principios de los años cincuenta. En 1952 investigadores de *Bell Labs* construyeron un sistema estadístico basado en correlación que podía reconocer con un 97-99 % de precisión cualquiera de los diez primeros números a partir de unos patrones grabados con los sonidos de las vocales de un único hablante (Davis, Biddulph y Balashek, 1952).

# Tema 3. Procesamiento del Lenguaje Natural

## Paradigma simbólico y estocástico: 1957-1970

A fines de la década de 1950 y comienzos de la década de 1960, el procesamiento del habla y el lenguaje se había dividido muy claramente en dos paradigmas:



Figura 2. Paradigmas a partir de la década de los 60. Fuente: elaboración propia.

### El paradigma simbólico

Apareció de dos líneas de investigación: la teoría del lenguaje formal y la inteligencia artificial.

La primera línea se basaba en el trabajo que realizaron Chomsky y sus colaboradores en la **teoría del lenguaje formal y la sintaxis generativa**, además de en el trabajo de muchos lingüistas e informáticos que estudiaban los algoritmos de análisis, inicialmente de arriba hacia abajo (*top-down*) y de abajo hacia arriba (*bottom-up*) y luego a través de la programación dinámica. Uno de los primeros sistemas de análisis fue TDAP (*Transformations and Discourse Analysis Project*) que implementó Zelig Harris entre junio de 1958 y julio de 1959 en la Universidad de Pennsylvania.

La segunda línea de investigación del paradigma simbólico fue el nuevo campo de la **inteligencia artificial**. En el verano de 1956 John McCarthy, Marvin Minsky, Claude Shannon y Nathaniel Rochester congregaron durante dos meses a un grupo de investigadores para realizar un simposio sobre lo que decidieron llamar "inteligencia artificial". Aunque el incipiente ámbito de la inteligencia artificial incluía una minoría de investigadores centrados en algoritmos estocásticos y estadísticos (incluidos los modelos probabilísticos y las redes neuronales), este nuevo campo se enfocó básicamente en el razonamiento y la lógica, representados por el trabajo de Newell y Simon en los programas de ordenador *Logic Theorist* y *General Problem Solver*.

# Tema 3. Procesamiento del Lenguaje Natural

(Newell, Shaw y Simon, 1959).

En los principios de la inteligencia artificial fue cuando se construyeron los primeros **sistemas de comprensión del lenguaje natural**. Estos sistemas simples estaban diseñados para trabajar en un dominio concreto y basaban su funcionamiento en la búsqueda de patrones y heurística de palabras clave para realizar el razonamiento y la búsqueda de respuestas. Fue a finales de la década de 1960 cuando se desarrollaron algunos sistemas lógicos más formales.

## El paradigma estocástico

Se desarrolló principalmente por parte de investigadores de los departamentos de estadística y de ingeniería electrónica. A fines de la década de 1950 comenzó a aplicarse el método bayesiano al problema del reconocimiento óptico de caracteres. Por ejemplo, Bledsoe y Browning construyeron un **sistema bayesiano** de reconocimiento de texto que calculaba la probabilidad de una secuencia de letras dadas las palabras de un diccionario.

En la década de 1960 aparecieron también los **primeros modelos psicológicos** para el PLN basados en gramáticas transformacionales y los primeros corpus disponibles *online*. Un ejemplo es el Brown Corpus, un corpus del inglés americano desarrollado en 1963 por la Brown University y que contenía una colección de un millón de palabras extraídas de 500 textos de diferentes géneros: periódicos, novelas, no ficción, académico, etc. (Kucera y Francis, 1967).

## Cuatro paradigmas de investigación: 1970-1983

En el siguiente período, en la década de 1970 y a principios de la década de 1980, se produce una explosión de la investigación en el procesamiento de lenguaje y del habla. Es en esta época cuando se desarrollan una serie de paradigmas de investigación que todavía hoy dominan el campo:

## Tema 3. Procesamiento del Lenguaje Natural



Figura 3. Paradigmas de investigación dominantes. Fuente: elaboración propia.

El **paradigma estocástico** jugó un papel muy importante en el desarrollo de algoritmos de reconocimiento de voz para los que se utilizaban modelos ocultos de Markov (HMM) y el teorema de codificación de canal de Shannon. Algunos de los investigadores que trabajaron en este ámbito fueron Jelinek, Bahl, Mercer y sus socios del Thomas J. Watson Research Center de IBM, Baker en la Carnegie Mellon University, e investigadores de los Bell Laboratories de AT&T.

El **paradigma basado en lógica** surgió del trabajo de Alain Colmerauer y sus colaboradores en la década de 1970 que desarrollaron Q-system, un analizador *bottom-up* basado en una serie de reglas con variables lógicas y que permitían la traducción del inglés al francés (Colmerauer, 1978). En este paradigma basado en lógica se engloba también la gramática de cláusulas definidas (*Definite Clause Grammar*, DCG), una forma de expresar la gramática en un lenguaje de programación lógico como por ejemplo Prolog (Pereira y Warren, 1980).

De forma independiente, apareció también el trabajo de Kay (1979) sobre la gramática funcional y, poco después, el trabajo de Joan Bresnan y Ronald Kaplan (1982) sobre la gramática léxico funcional (*Lexical functional grammar*, LFG), una gramática generativa que se centra en la investigación de la sintaxis del lenguaje natural.

El **paradigma de la comprensión del lenguaje natural** apareció durante la década de 1970 con la aparición del sistema SHRDLU (Winograd, 1972). Este sistema simulaba un robot que movía bloques y era capaz de recibir comandos del lenguaje natural en formato de texto, como por ejemplo “mueve el bloque rojo que se encuentra en la parte superior del verde más pequeño”. Este sistema, complejo y

## Tema 3. Procesamiento del Lenguaje Natural

sofisticado para su época, también fue el primero en construir una gramática relativamente extensa del inglés.

Los avances en el modelo de análisis del lenguaje de Winograd dejó claro que la investigación debía comenzar a enfocarse en la semántica y los modelos de discurso. Roger Schank y sus colaboradores, conocidos como la Escuela de Yale, construyeron una serie de programas de comprensión del lenguaje que se centraron en el conocimiento humano, por ejemplo, en planes y objetivos, y la organización de la memoria humana (Schank y Riesbeck, 1981). Estos trabajos, por ejemplo el realizado por R. F. Simmons (1973), usaban una representación del conocimiento en forma de red, lo que se conoce como redes semánticas (Quillian, 1968), y comenzaron a incorporar en sus representaciones la idea de gramática de casos (Fillmore, 1968), por la cual se establece una relación entre un verbo y múltiples papeles temáticos que serían los sintagmas nominales.

El paradigma basado en lógica y el paradigma de la comprensión del lenguaje natural se unificaron en sistemas que usaban la **lógica de predicados** como una representación semántica, como el sistema de búsqueda de respuestas LUNAR (Woods, 1967).

El **paradigma del modelado del discurso** se centró en las cuatro áreas clave del discurso. Grosz y sus colaboradores introdujeron el estudio de la estructura del discurso y el enfoque del discurso (Grosz, 1977) (Grosz y Sidner, 1986). Una serie de investigadores comenzaron a trabajar en la resolución de forma automática de referencias en el discurso (Hobbs, 1978) (Hobbs, 1979). Por último, se desarrolló el modelo BDI (creencias-deseos-intenciones en inglés), un marco de trabajo basado en la lógica de actos de habla (Perrault y Allen, 1980) (Cohen y Perrault, 1979).

### Revivir del empirismo y los modelos de estados finitos: 1983-1993

A partir de 1983 volvieron dos clases de modelos que habían perdido popularidad a finales de la década de 1950 y principios de la década de 1960 debido a los

# Tema 3. Procesamiento del Lenguaje Natural

argumentos teóricos en su contra (Chomsky, 1959).

## Tema 3. Procesamiento del Lenguaje Natural

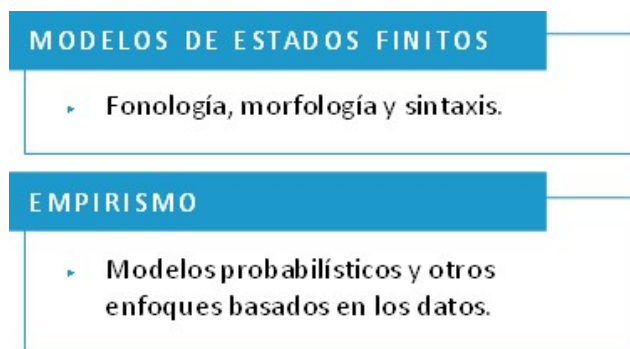


Figura 4. Modelos de estados finitos y modelos empíricos. Fuente: elaboración propia.

Los **modelos de estados finitos** revivieron en esta época y comenzaron a recibir atención porque se usaron en la fonología y la morfología (Kaplan y Kay, 1981), y en la sintaxis (Church, 1980).

La segunda tendencia en este período fue lo que se ha llamado el **retorno del empirismo**, marcada por el aumento de los modelos probabilísticos para el procesamiento del lenguaje y del habla. Cabe destacar la influencia del trabajo de los investigadores del Thomas J. Watson Research Center de IBM sobre modelos probabilísticos en el reconocimiento de voz. Estos métodos probabilísticos y otros enfoques basados en los datos se extendieron al etiquetado morfosintáctico (POS *tagging*), al análisis y resolución de ambigüedades y a la semántica.

Este paradigma empírico vino acompañado por el enfoque de la evaluación de los modelos basado en los datos. Entonces en este período se desarrollaron métricas cuantitativas para la evaluación y se enfatizó en la comparación del rendimiento de estas métricas con los resultados de las investigaciones previas. Además, en este período, se trabajó considerablemente en la generación de lenguaje natural.

### Unión de las diferentes vertientes: 1994-1999

En los últimos cinco años del pasado milenio el campo del procesamiento del lenguaje natural sufrió grandes cambios.

## Tema 3. Procesamiento del Lenguaje Natural

En primer lugar, los modelos probabilísticos y los modelos basados en datos se volvieron estándares para el procesamiento del lenguaje natural. Los algoritmos de análisis, de etiquetado morfosintáctico, de resolución de referencias y de procesamiento del discurso empezaron a incorporar probabilidades y adoptaron metodologías de evaluación provenientes de los ámbitos del **reconocimiento de la voz y la recuperación de información**.

En segundo lugar, el aumento en la velocidad y la memoria de los ordenadores permitió la explotación comercial de varias áreas del procesamiento del lenguaje y del habla, en particular el reconocimiento de la voz y la revisión de la ortografía y la gramática. Además, los algoritmos de procesamiento del lenguaje y del habla comenzaron a aplicarse a la comunicación aumentativa para ayudar a personas con algún tipo de discapacidad. Por último, el aumento de la web enfatizó la necesidad de la recuperación y la extracción de información basada en el lenguaje natural.

### Auge del aprendizaje automático: 2000

Las tendencias empiristas que marcaron la última parte de la década de 1990 se aceleraron a un ritmo asombroso en el nuevo milenio. Esta aceleración fue impulsada en gran parte por el auge del **aprendizaje automático**.

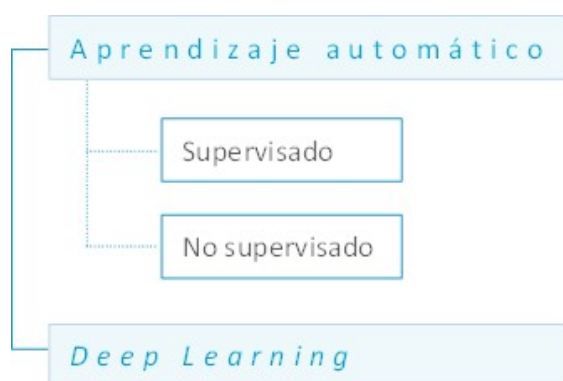


Figura 5. Aprendizaje automático o machine *learning*. Fuente: elaboración propia.

Grandes cantidades de material hablado y escrito se pusieron a disposición de los

## Tema 3. Procesamiento del Lenguaje Natural

investigadores a través de organizaciones tipo el *Linguistic Data Consortium* (LDC). Estos materiales eran fuentes de texto estándar que venían anotados con información sintáctica, semántica y pragmática. Entre estos materiales caben destacar las primeras colecciones anotadas: el *Penn Treebank*, el *Prague Dependency Treebank*, que anota la estructura de las dependencias, y las anotaciones semánticas del *PropBank*.

La existencia de estos recursos anotados promovió la tendencia de atacar los problemas más complejos del procesamiento del lenguaje natural, tipo el análisis sintáctico y semántico, como problemas de **aprendizaje automático supervisado**.

Por ejemplo, se empezaron a aplicar las máquinas de vectores de soporte (SVM), el principio de máxima entropía, la regresión logística multinomial y los modelos bayesianos en la lingüística computacional. Entonces, este mayor enfoque en el aprendizaje automático llevó a una interacción más seria de los lingüistas computacionales con la comunidad estadística.

El coste y la dificultad de producir corpus anotados se convirtió en un factor limitante del uso de los enfoques supervisados para muchos problemas del procesamiento del lenguaje. Es por eso por lo que a partir de 2005 aparece una nueva tendencia hacia el uso de técnicas de **aprendizaje no supervisado** en el procesamiento del lenguaje natural. Entonces se empezaron a construir algunas aplicaciones lingüísticas a partir de datos sin anotación alguna, por ejemplo, para la traducción automática o para el modelado de temas.

Los algoritmos de aprendizaje no supervisado se han usado en el **etiquetado morfosintáctico** (*POS tagging*) para agrupar palabras en las correspondientes partes del lenguaje (Goldwater y Griffiths, 2007) (Sirts, Eisenstein, Elsnér y Goldwater, 2014). Además, las técnicas del aprendizaje no supervisado se han usado también para el **etiquetado semántico**, donde se han creado conjuntos de roles semánticos a partir de las características sintácticas (Titov y Klementiev, 2012)

## Tema 3. Procesamiento del Lenguaje Natural

(Lang y Lapata, 2014).

En 2006 Geoffrey Hinton acuña el término **Deep Learning** (aprendizaje profundo). Con el auge de este tipo de redes neuronales en la década de 2010, estas redes de neuronas artificiales profundas se empezaron a usar en diferentes ámbitos del procesamiento del lenguaje natural. Las **redes neuronales recurrentes** son ya una realidad que se está utilizando como una alternativa a los modelos ocultos de Markov (HMM) en **análisis morfosintáctico y en el análisis sintáctico** (Chen y Manning, 2014) (Dozat, Qi, y Manning, 2017). Además, las redes neuronales profundas también se están utilizando para el **etiquetado semántico** (Collobert et al., 2011) (Foland Jr. y Martin, 2015). De hecho, el *Deep learning* es la base de los modelos de secuencia a secuencia (seq2seq) que se utilizan en los agentes conversacionales y *chatbots* actuales.

# Tema 3. Procesamiento del Lenguaje Natural

## 3.3. Conocimiento del lenguaje utilizado en el PLN

Lo que distingue a las aplicaciones de procesamiento de lenguaje de otros sistemas de procesamiento de datos es su **uso del conocimiento del lenguaje**. Por ejemplo, un programa que cuenta el número de bytes, palabras y líneas en un archivo de texto podemos considerarlo como una **aplicación de procesamiento de datos ordinaria**. Sin embargo, si para contar las palabras en el archivo de texto se requiere conocimiento sobre lo que significa ser una palabra porque se quieren contar el número de pronombres que aparecen en el archivo, ese programa se convierte en un **sistema de procesamiento de lenguaje natural**.

Por supuesto, este sistema de procesamiento de lenguaje es extremadamente simple en comparación con los agentes conversacionales, los sistemas de traducción automática y los sistemas de búsqueda de respuestas, que requieren un conocimiento mucho más amplio y profundo del lenguaje.



Figura 6. Conocimientos necesarios para tareas complejas del PLN. Fuente: elaboración propia.

### Fonética y fonología: conocimiento sobre los sonidos lingüísticos

Un agente conversacional necesita reconocer las palabras de una señal de audio y generar una señal de audio de una secuencia de palabras. Estas tareas de reconocimiento de la voz y del habla son tareas de síntesis que requieren conocimiento sobre **fonética** y **fonología**. Es decir, conocimiento de cómo se

## Tema 3. Procesamiento del Lenguaje Natural

pronuncian las palabras a partir de la secuencia de sonidos y cómo se generan cada uno de estos sonidos acústicamente.

### Morfología: conocimiento de los componentes significativos de las palabras

El agente conversacional también necesita reconocer y saber producir variaciones de las palabras, por ejemplo, reconocer que “puertas” es el plural de una palabra y saber generar una frase en la que esta palabra se utilice en plural. Entonces, estas tareas requieren conocimiento sobre **morfología**, es decir, la forma en que las palabras se descomponen en partes que tienen un significado como puede ser la raíz de la palabra y una terminación que indique que es un plural.

### Sintaxis: conocimiento de las relaciones estructurales entre palabras

Si vamos más allá de las palabras como elementos aislados y entendidas de forma individual, un agente conversacional debe usar conocimiento estructural para encadenar las palabras que constituyen su respuesta. El agente debe saber identificar que una secuencia de palabras no tiene sentido a pesar de que el conjunto de palabras original pudiera tener sentido si se ordenaran las palabras de otra forma. El conocimiento necesario para ordenar y agrupar palabras se llama **sintaxis**.

### Semántica: conocimiento del significado

Para responder a una pregunta, el agente conversacional puede necesitar saber algo sobre la **semántica léxica**, es decir, sobre el significado de cada una de las palabras, así como sobre **semántica composicional** o el significado de varias palabras que se utilizan de forma conjunta en una combinación de palabras.

Además de comprender el significado de las palabras, el agente conversacional puede necesitar saber algo sobre la relación de las palabras con la estructura sintáctica: si un sintagma es un complemento circunstancial de tiempo o un complemento del nombre. Por lo tanto, el conocimiento sobre la sintaxis y el conocimiento sobre la semántica se van a utilizar de forma conjunta en el procesamiento del lenguaje natural.