

# conexionismo – connectionism

Authored by  
**memjavad**

November 21, 2025

## RECOMMENDED CITATION

memjavad (2025). *conexionismo – connectionism*. Spanish Psychological Databases.  
Retrieved from <https://spanish.arabpsychology.com/?p=5650>

## Conexionismo

**Primary Disciplinary Field(s):** Ciencia Cognitiva, Inteligencia Artificial (IA), Neurociencia Computacional, Psicología.

### 1. Definición Central y Paradigma

El conexionismo, también conocido como procesamiento distribuido en paralelo (PDP), representa un enfoque teórico fundamental dentro de la **ciencia cognitiva**, la **inteligencia artificial** y la **neurociencia computacional**. Este paradigma busca modelar los fenómenos mentales y conductuales a través de sistemas que se asemejan a redes neuronales biológicas. A diferencia del cognitivismo clásico, que postula que la cognición opera mediante la manipulación de símbolos discretos y reglas sintácticas (la metáfora de la mente como software), el conexionismo propone que el conocimiento no reside en reglas explícitas o símbolos localizados, sino en los patrones de **conexiones ponderadas** entre una vasta cantidad de unidades de procesamiento simples, interconectadas en una red.

La característica definitoria del conexionismo es su énfasis en el procesamiento distribuido. La información, en lugar de estar almacenada en una ubicación específica, se dispersa a través de la intensidad de las conexiones sinápticas (los pesos) de la red. Esto confiere a los modelos conexionistas propiedades clave como la robustez, la tolerancia a fallos y la capacidad inherente de **aprendizaje** y generalización. El aprendizaje se logra típicamente mediante algoritmos que modifican sistemáticamente estos pesos de conexión basándose en la experiencia, permitiendo que la red se ajuste gradualmente para mapear entradas deseadas a salidas correctas. Este proceso de ajuste de pesos es fundamentalmente lo que permite a las redes neuronales adquirir habilidades cognitivas complejas, desde el reconocimiento de patrones visuales hasta la comprensión del lenguaje, diferenciándose de la programación explícita requerida por los sistemas simbólicos tradicionales.

Este marco teórico ofrece una alternativa poderosa y biológicamente plausible al modelo simbólico tradicional, especialmente al abordar la naturaleza intrínsecamente ruidosa, incompleta o ambigua de los datos sensoriales del mundo real. Mientras que los modelos clásicos luchaban por explicar cómo los sistemas cognitivos manejan tales desafíos, las redes conexionistas demuestran una notable capacidad para la **recuperación asociativa** y la inferencia contextual. La potencia del enfoque radica en que el sistema global exhibe una inteligencia compleja que emerge de la interacción masiva y simple de sus componentes individuales. El conexionismo ha evolucionado enormemente desde sus formulaciones iniciales, siendo hoy la base teórica de las tecnologías de **aprendizaje profundo** (Deep Learning) que dominan la inteligencia artificial contemporánea, confirmando su validez como un marco unificador para entender tanto la computación biológica como la artificial.

## 2. Orígenes y Desarrollo Histórico

Los orígenes del pensamiento conexionista se remontan a mediados del siglo XX, mucho antes de que el término se popularizara. La primera formalización matemática de una unidad de procesamiento neuronal data de 1943, con el trabajo pionero de Warren McCulloch y Walter Pitts. Ellos propusieron un modelo de neurona artificial que podía realizar funciones lógicas simples, sentando las bases para el concepto de **computación neuronal**. Sin embargo, el verdadero impulso fundacional provino de Donald Hebb en 1949, quien en su obra *The Organization of Behavior* postuló la famosa regla de aprendizaje: «Las neuronas que se disparan juntas, se conectan juntas» (Regla de Hebb). Esta regla proporcionó el primer mecanismo plausible para que las redes neuronales modificaran sus conexiones basándose en la actividad correlacionada, vinculando el aprendizaje con la plasticidad sináptica y sugiriendo que la memoria se almacena en el fortalecimiento de las sinapsis.

El desarrollo inicial fue seguido por la creación de modelos prácticos, siendo el más destacado el Perceptrón, introducido por Frank Rosenblatt en 1957. El Perceptrón fue la primera red neuronal capaz de aprender a clasificar patrones mediante un algoritmo de corrección de errores, generando gran entusiasmo sobre el potencial de las máquinas para aprender. No obstante, el campo sufrió un revés significativo en 1969 con la publicación del libro *Perceptrons* de Marvin Minsky y Seymour Papert. Este trabajo demostró rigurosamente las limitaciones de los Perceptrones de una sola capa, probando que no podían resolver problemas no linealmente separables, como el fundamental problema de la **OR exclusiva (XOR)**. Esta crítica, que implicó que las redes simples no podían modelar funciones lógicas básicas, junto con el dominio del paradigma simbólico en la IA, condujo al primer «Invierno de la IA» y marginó la investigación conexionista durante más de una década, enfocando la atención de la comunidad científica en los sistemas expertos y la lógica formal.

El resurgimiento, o lo que se conoce como la «segunda ola» del conexionismo, ocurrió a principios de la década de 1980. El factor crucial fue el redescubrimiento y la aplicación exitosa del algoritmo de **retropropagación** (backpropagation), formalizado por David Rumelhart, Geoffrey Hinton y Ronald Williams en 1986. Este algoritmo permitió entrenar eficientemente redes multicapa (redes con capas ocultas), superando las limitaciones del Perceptrón al permitir que las redes aprendieran representaciones internas complejas y resolvieran problemas no lineales. La publicación del influyente volumen *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (1986) por el Grupo PDP (Rumelhart, McClelland y otros) consolidó el conexionismo como un rival legítimo y poderoso del cognitivismo clásico. Este trabajo no solo proporcionó las herramientas algorítmicas, sino también un marco teórico coherente que proponía que la mente era inherentemente paralela y distribuida, marcando el inicio de su dominio en la IA moderna que culminó con la revolución del aprendizaje profundo.

### 3. Principios Fundamentales del Procesamiento Distribuido

El paradigma conexionista se sustenta en varios principios que lo distinguen fundamentalmente de la computación simbólica tradicional. El más esencial es el principio de que la cognición es el resultado de la interacción dinámica de un gran número de unidades simples, donde la inteligencia no es programada explícitamente, sino que **emerge** de la interacción colectiva. Cada unidad de procesamiento (neurona artificial) recibe entradas, realiza una suma ponderada y aplica una función de activación (a menudo no lineal) para producir una salida. Esta simplicidad a nivel de unidad contrasta con la complejidad de las estructuras de datos y algoritmos en los modelos simbólicos, pero permite que la red, en su conjunto, realice transformaciones de datos altamente sofisticadas.

Otro principio clave es la **representación distribuida**. En lugar de que un concepto (por ejemplo, "perro") esté codificado por un único nodo o símbolo (representación localista), la información sobre el perro se distribuye a través de un patrón de activación en muchas unidades. Este patrón distribuido no solo proporciona robustez --si algunas unidades fallan o están ruidosas, el concepto aún puede ser recuperado por las unidades restantes--, sino que también permite la **superposición de representaciones**. Diferentes conceptos pueden compartir unidades de procesamiento, lo que facilita la generalización y la transferencia de conocimiento entre tareas relacionadas. Por ejemplo, las características compartidas entre "perro" y "lobo" se codificarían en patrones de activación parcialmente superpuestos, permitiendo inferencias rápidas sobre el otro animal.

El aprendizaje en el conexionismo es fundamentalmente **inductivo y basado en datos**. A diferencia de los sistemas expertos que requieren la codificación manual de reglas por parte de humanos, las redes conexionistas aprenden automáticamente a partir de ejemplos y mediante la corrección de errores. Este proceso de aprendizaje se rige por la modificación iterativa de los pesos sinápticos, siguiendo reglas como la regla de Hebb o, más comúnmente, el algoritmo de gradiente descendente (como la retropropagación). La capacidad de aprender directamente de la experiencia permite a estos sistemas adaptarse a entornos cambiantes y descubrir estructuras latentes en los datos que son demasiado complejos o sutiles para ser identificados y codificados por los programadores humanos, lo que explica su éxito en tareas de percepción y clasificación.

**Paralelismo Masivo:** Las operaciones se realizan simultáneamente a través de miles o millones de unidades, reflejando la arquitectura del cerebro biológico y permitiendo un procesamiento increíblemente rápido y eficiente de grandes conjuntos de datos, esencial para el rendimiento en tiempo real.

**Almacenamiento en Pesos:** El conocimiento del sistema (la memoria a largo plazo) no se almacena como datos en registros, sino como la fuerza de las conexiones entre las unidades. La memoria de trabajo, en contraste, se considera el patrón de activación temporal de las unidades.

**Generalización:** Una vez entrenadas, las redes conexionistas tienen la capacidad intrínseca de manejar entradas nuevas o ligeramente modificadas (que no vieron durante el entrenamiento) y producir salidas coherentes, lo cual es crucial para la inteligencia real y la adaptabilidad.

**Tolerancia a Fallos (Gracia de la Degradación):** La destrucción parcial de la red resulta en una degradación gradual del rendimiento en lugar de un colapso catastrófico. Esta robustez es una consecuencia directa de la naturaleza distribuida y redundante del almacenamiento de la información.

#### 4. Modelos Clave y Arquitecturas

El conexionismo abarca una amplia gama de arquitecturas de redes neuronales, cada una diseñada para abordar tipos específicos de problemas computacionales y cognitivos. Históricamente, el modelo más influyente en la cognición fue la red de **Perceptrón Multicapa (MLP)** entrenada con retropropagación. Estas redes, caracterizadas por una capa de entrada, una o más capas ocultas y una capa de salida, fueron esenciales para demostrar que las redes podían aprender mapeos complejos y no lineales, sirviendo como el caballo de batalla para la simulación de tareas cognitivas básicas, como la conjugación de verbos irregulares, la clasificación de patrones y el modelado de la memoria asociativa. Las MLP establecieron la viabilidad del aprendizaje profundo al introducir la idea de que capas sucesivas podían aprender representaciones de características de complejidad creciente.

Con el avance de la potencia de cálculo y el volumen de datos, surgieron arquitecturas más especializadas para manejar secuencias temporales. Las **Redes Neuronales Recurrentes (RNNs)** son fundamentales para el procesamiento de secuencias, ya que contienen bucles de retroalimentación que permiten que la información persista y que el estado actual de la red dependa de entradas previas, dándoles una forma de "memoria" contextual. Esto las hace ideales para tareas como el modelado del lenguaje y el reconocimiento de voz. Sin embargo, las RNNs básicas a menudo sufren del problema del gradiente desvaneciente (vanishing gradient), lo que dificulta el aprendizaje de dependencias a largo plazo. Esto llevó al desarrollo de variantes más sofisticadas como las redes **LSTM** (Long Short-Term Memory) y **GRU** (Gated Recurrent Units), que utilizan mecanismos de "puertas" (gates) para controlar el flujo de información y mantener la relevancia contextual a lo largo de secuencias extensas.

Quizás la arquitectura más transformadora en la última década ha sido la **Red Neuronal Convolutiva (CNN)**. Inspiradas en la organización jerárquica del córtex visual, las CNNs utilizan capas de convolución que aplican filtros a pequeñas regiones de la entrada (especialmente imágenes), permitiendo que la red aprenda jerarquías de características espaciales, desde bordes simples hasta objetos complejos y caras. Las CNNs han revolucionado la visión por computadora y el procesamiento de imágenes médicas. Más recientemente, la introducción de la arquitectura **Transformer** ha desplazado a las RNNs en muchas tareas de

procesamiento del lenguaje natural (PLN). Los Transformers, que dependen del mecanismo de auto-atención (self-attention), permiten a la red evaluar la importancia relativa de diferentes partes de la entrada de manera paralela, permitiendo un procesamiento altamente eficiente y una comprensión contextual del lenguaje sin precedentes, lo que ha generado los modelos fundacionales de lenguaje a gran escala (LLMs).

**Redes Hopfield:** Modelos tempranos utilizados para la memoria asociativa y la optimización, caracterizados por ser redes recurrentes simétricas que convergen a estados de energía mínimos, actuando como atractores de memoria para recuperar patrones incompletos o ruidosos.

**Máquinas de Boltzmann y Redes de Creencia Profunda (DBNs):** Modelos probabilísticos utilizados para el aprendizaje no supervisado y generativo, fundamentales en los primeros intentos de aprendizaje profundo y en la extracción de características no etiquetadas.

**Redes Neuronales Convolucionales (CNNs):** Arquitecturas especializadas para datos con estructura de cuadrícula, utilizando pesos compartidos y capas de *pooling* para lograr invarianza traslacional, crucial en el reconocimiento de objetos independientemente de su ubicación exacta.

**Redes Adversarias Generativas (GANs):** Compuestas por dos redes que compiten (un generador y un discriminador), utilizadas para crear contenido sintético altamente realista, demostrando la capacidad del conexionismo para modelar distribuciones de datos complejas y generar nuevas instancias.

## 5. Aplicaciones en la Ciencia Cognitiva y la Inteligencia Artificial

El impacto del conexionismo en la **ciencia cognitiva** es profundo, ofreciendo un marco robusto para simular y comprender procesos mentales que el cognitivismo clásico encontraba difíciles de explicar. Los modelos conexionistas han sido utilizados con éxito para abordar problemas clásicos de la psicología, como la adquisición del lenguaje, particularmente el aprendizaje de la morfología y la fonología, la formación de conceptos, la toma de decisiones y la memoria. Por ejemplo, los modelos PDP han demostrado cómo la regularización de verbos (como "caminar-caminé") y la posterior sobre-regularización de verbos irregulares ("hacer-hicó" en niños) pueden emerger naturalmente de un sistema de aprendizaje basado en patrones y frecuencias de entrada, sin necesidad de postular reglas gramaticales innatas explícitas, lo que ofrece una perspectiva de desarrollo más gradual y basada en la experiencia.

En el campo de la **Inteligencia Artificial (IA)**, el conexionismo, bajo su forma moderna de aprendizaje profundo, constituye el paradigma dominante que ha impulsado la reciente explosión de la IA. Las aplicaciones son vastas y han transformado industrias enteras. Los sistemas basados en redes neuronales profundas (DNNs) son responsables de los avances en el reconocimiento de imágenes, donde superan consistentemente el rendimiento humano en tareas de clasificación y detección. En el procesamiento del lenguaje natural (PLN), modelos basados en Transformers como GPT-4 y BERT han logrado capacidades de comprensión, generación y

traducción que antes se consideraban inalcanzables, permitiendo la creación de asistentes virtuales avanzados, sistemas de diálogo coherentes y herramientas de resumen automático de alta calidad.

Adicionalmente, el conexionismo ha fortalecido la **neurociencia computacional** al proporcionar herramientas matemáticas para modelar la función cerebral. Al construir modelos que imitan la estructura y función del cerebro (como las redes recurrentes que modelan circuitos corticales o las CNNs que replican el procesamiento visual), los investigadores pueden generar hipótesis comprobables sobre cómo el cerebro real codifica, almacena y procesa la información. Estos modelos no solo son herramientas de ingeniería, sino también herramientas científicas que permiten la experimentación virtual y la comprensión de patologías cognitivas, como la dislexia o ciertos tipos de amnesia, al simular la degradación o el daño a las conexiones de la red, ofreciendo una visión mecanicista de los déficits cognitivos.

## 6. Debates Filosóficos y Críticas al Conexionismo

A pesar de su éxito práctico y su creciente plausibilidad biológica, el conexionismo ha sido objeto de intensos **debates filosóficos**, principalmente desde la perspectiva del cognitivismo clásico y la filosofía de la mente. La crítica más famosa es la formulada por Jerry Fodor y Zenon Pylyshyn en 1988, conocida como la crítica de la **composición y la sistematicidad**. Argumentan que la cognición humana es inherentemente sistemática (si un humano puede entender "Juan ama a María", puede entender "María ama a Juan") y composicional (el significado de una oración se compone del significado de sus partes). Fodor y Pylyshyn sostienen que los modelos conexionistas tradicionales tienen dificultades para manejar estas propiedades, ya que sus representaciones distribuidas no siempre permiten la manipulación explícita de constituyentes simbólicos discretos, lo que sugiere que el conexionismo podría ser una teoría de la implementación más que una teoría de la naturaleza fundamental de la cognición.

Otra crítica significativa se centra en el problema de la **transparencia y la interpretabilidad**, exacerbado por el aprendizaje profundo. Los sistemas conexionistas, especialmente las redes profundas con millones de parámetros, a menudo funcionan como "cajas negras". Es extremadamente difícil para los investigadores rastrear el camino causal de una entrada a una salida o entender por qué la red tomó una decisión particular, ya que el conocimiento está codificado de manera implícita y distribuida. Esta falta de interpretabilidad es un problema serio no solo para la comprensión científica de la cognición, sino también para la aplicación práctica en campos críticos como la medicina, las finanzas o la justicia, donde se requiere la explicabilidad y la rendición de cuentas de las decisiones de la IA. Aunque el campo de la IA Explicable (XAI) está trabajando para mitigar esto, la opacidad sigue siendo una limitación inherente a los modelos basados en la emergencia de patrones.

Finalmente, existe el debate sobre la **plausibilidad biológica versus la eficiencia computacional**. Si bien los modelos conexionistas se inspiraron inicialmente en el cerebro, muchos de los algoritmos más exitosos de la IA (como la retropropagación en su forma pura) tienen una baja plausibilidad biológica, ya que el cerebro no parece utilizar un mecanismo de retropropagación idéntico y simétrico. Esto ha llevado a una división: algunos investigadores se centran en modelos estrictamente bioinspirados, buscando mecanismos de aprendizaje más plausibles como el aprendizaje hebbiano o la propagación de errores hacia adelante (neurociencia computacional), mientras que otros priorizan la eficiencia y el rendimiento (aprendizaje profundo). Esta divergencia plantea la pregunta crítica de si los modelos de IA más avanzados realmente nos están enseñando algo fundamental sobre la arquitectura de la mente humana, o si son simplemente métodos de optimización matemática altamente efectivos para resolver problemas de ingeniería.

## 7. Lecturas Adicionales

[Connectionism - Wikipedia](#)

[Connectionism \(Stanford Encyclopedia of Philosophy\)](#)

[Parallel Distributed Processing: Explorations in the Microstructure of Cognition \(Rumelhart, McClelland, et al.\)](#)

[The Deep Learning Revolution \(Nature insight\)](#)