

El conexionismo como marco de simulación: Aplicación a una tarea de facilitación semántica

ALFONSO PITARQUE y SALVADOR ALGARABEL
Universidad de Valencia

Resumen

El conexionismo es un método de simulación de la cognición basado en la interacción en paralelo de redes de unidades (de inspiración neuronal). Este trabajo presentará algunas de las características básicas que configuran lo que hoy entendemos por conexionismo (punto 1), así como los elementos básicos que desde un punto de vista formal configuran una arquitectura conexionista (punto 2) para después presentar una simulación sobre el conocido efecto de facilitación semántica (semantic priming; punto 3) que pretende demostrar la validez psicológica de dichos modelos al ser capaces de dar fe de los datos experimentales hasta hoy considerados en exclusividad por los modelos localistas y simbólicos de memoria semántica (p. e., Collins y Loftus, 1975; Anderson, 1983, entre otros muchos). Un último punto llevará a cabo una breve valoración crítica del enfoque conexionista.

Palabras clave: conexionismo, procesamiento distribuido en paralelo (PDP), redes neurales, representaciones localistas vs. distribuidas, asociador de patrones, efecto de facilitación semántica (priming).

CONNECTIONISM AS A MODELING METHOD: APPLICATION TO A SEMANTIC PRIMING TASK

Abstract

Connectionism is a method of modeling cognition as the interaction of neuron-like units. In this article we review the theoretical impact and general principles (part 1) and formal assumptions (part 2) of this framework. One simulation (part 3) will try to test its psychological validity in relation to the assumptions made by the symbolic and localist models of semantic memory (Collins y Loftus, 1975; Anderson, 1983 and others). A brief and critical valuation of the connectionism view will finish our article.

Key words: connectionism, parallel distributed processing (PDP), neural networks, local vs distributed processing, pattern associator, semantic priming effect.

Agradecimientos: Esta investigación fue subvencionada por la beca PB88-0496 de Promoción General del Conocimiento de la Dirección General de Investigación Científica y Técnica del M.E.C. Queremos agradecer desde aquí los comentarios y sugerencias hechas por los dos revisores anónimos de este artículo.

Dirección de los autores: Facultad de Psicología. Universidad de Valencia. Avda. Blasco Ibáñez. 46010 Valencia.

INTRODUCCION

Dada la importancia actual del conexionismo como marco paradigmático, y como conjunto de instrumentos de simulación, el presente artículo lleva a cabo una revisión crítica de sus principios y su espectro de acción dentro del campo de la psicología. Con este fin, este artículo presenta primeramente las características generales del conexionismo, para después analizar los componentes básicos de un modelo conexionista, y finalmente presentar un programa conexionista de simulación del conocido efecto de «facilitación semántica» (semantic priming; según el cual el procesamiento de un concepto se ve facilitado por el procesamiento anterior de un concepto semánticamente relacionado con él; véase, p. e., Meyer y Schvaneveldt, 1971) hallados ante un amplio de tareas cronométricas (decisión léxica, naming, etc.). Dado que tal efecto experimental ha sido usualmente tomado como la evidencia experimental básica en favor de los llamados modelos de representación localicista y simbólica del conocimiento (p. e., Collins y Loftus, 1975; Anderson, 1983, entre otros muchos), una simulación adecuada de dichos resultados hablaría en favor así mismo de la validez psicológica de los modelos conexionistas. Una breve valoración crítica del enfoque que nos ocupa cerrará nuestro trabajo.

1. CARACTERÍSTICAS GENERALES DEL ENFOQUE CONEXIONISTA

El conexionismo (Feldman, 1981) es un concepto extenso por el que nos referimos tanto a una forma de «hacer teoría» como, y en consecuencia, a un amplio conjunto de instrumentos metodológicos orientados a la simulación de la cognición humana. El conexionismo se ha ido construyendo a partir de las aportaciones de diversos campos científicos (ingeniería, informática, psicología, etc.) que están plenamente integrados en lo que modernamente se entiende como «Ciencia Cognitiva» (Pylyshyn, 1984).

El denominador común a todos los modelos conexionistas es la admisión de que el procesamiento acaece sobre poblaciones amplias de *unidades* de procesamiento muy simples (de inspiración pseudoneuronal), interactuantes entre sí a través de las *conexiones* que las asocian. Toda unidad de la red viene definida por un *nivel de activación* expresado de forma cuantitativa, de igual modo que toda conexión viene caracterizada por un valor de fuerza del trazo o *peso* de la conexión, también expresado de forma cuantitativa. La misión de cada unidad es simplemente computar la activación recibida de sus unidades vecinas en base a los pesos de las conexiones que las asocian y, a través de determinada función aplicada sobre tal input, computar el output que envía a las unidades con las que se halla relacionada. Tal interacción producirá asimismo la modificación constante de las conexiones de la red (véase, p. e., Hinto, en prensa; Massaro, 1989; McClelland, 1988; o Smolensky, 1988, para revisiones teóricas). El conexionismo asume pues que la cognición acaece vía la continua interacción *en paralelo* de tales elementos, frente al procesamiento serial propugnado por los modelos más clásicos de simulación de la cognición humana. Se inscribiría pues dentro del área de la computación conocida genéricamente como «redes neurales», ha-

biéndosele bautizado también como *enfoque PDP* (de Pararell Distributed Processing en relación al título de los volúmenes de McClelland, Rumelhart and the PDP group, 1986 y Rumelhart, McClelland and the PDP group, 1986). Más abajo entraremos en una consideración más detallada de los aspectos formales involucrados en toda arquitectura conexionista.

Pese a la apariencia nueva de este enfoque, el conexionismo ha tenido antecedentes claros a lo largo de la historia en las distintas áreas de investigación sobre la cognición humana. Podríamos asumir que los orígenes conceptuales más remotos del conexionismo cabría buscarlos en las aportaciones de neurofisiólogos como Hebb (1949), Lashley (1950) y Luria (1966). Así mientras que Hebb formuló su regla de modificación de las sinapsis neurales y su idea de la «asamblea neuronal» como unidad básica de procesamiento, Lashley, por su parte, con su «principio de equipotencialidad» y la incapacidad de localización del engrama anticipó claramente la idea de una representación distribuida del conocimiento. De igual modo su idea de «sistema funcional dinámico» expresaba que toda conducta cognitiva resulta de la coordinación de un amplio número de componentes elementales, localizados en distintas partes del cerebro, pero interactuantes en su acción.

El modelo de neurona formal fue introducido en los años 40 por McCulloch y Pitts (1943) como modelo general de la computación. De cualquier modo habrá que esperar a los primeros desarrollos informáticos (mediados de los años 50 y principios de los 60) para que tal idea pueda comenzar a ser implementada en un lenguaje de ordenador. En tal sentido destacan como pioneras las aportaciones de Rosenblatt (1962), Selfridge (1959) y Widrow y Hoff (1960). El primero articuló el modelo del «perceptrón» como sistema interactivo de clasificador de patrones, formulado en forma de red conexionista sin unidades ocultas y con capacidad de aprendizaje. Por su parte Selfridge en su «pandemonium» defendió planteamientos similares trasladados al campo de la percepción. Widrow y Hoff participaron en el desarrollo del procedimiento llamado de la «media cuadrática menor», más tarde conocida como regla delta de aprendizaje, que desempeña un papel importante entre los instrumentos metodológicos de simulación (véase apartado 2 para una ampliación sobre el significado de tales conceptos).

Si la aportación de Rosenblatt supone el nacimiento del esquema conexionista, la crítica de Minsky y Papert (1969) a tales trabajos supondrá su muerte temporal (Schenider, 1987). Dicho trabajo mostró como tales programas sólo eran capaces de aprender tareas lógicas simples. Las soluciones a estos problemas sólo quedará definitivamente resuelta hasta bien entrados los años 80, con el desarrollo de algoritmos de aprendizaje y modelos que superen las limitaciones previstas por Minsky y Papert.

Mientras que los inicios de la década de los 70 estuvo dominada por los modelos que propugnaron el procesamiento serial de la información (p. e., J. R. Anderson, 1976; Anderson y Bower, 1973), a partir de la mitad de los años 70 se comienza a abordar firmemente el tema del procesamiento distribuido y en paralelo de la información. Aportaciones pioneras como el manual de Kohonen (1977) planteando los principios básicos de una arquitectura asociativa; las de Grossberg (1976), en torno al análisis matemático de las propiedades de los modelos neurales y el aprendizaje; J. A. Anderson (1973, 1977), sobre la relevancia de las representaciones distribuidas; Marr y Poggio (1976) sobre un modelo computacional de percepción estereoscópica; los

modelos de lectura de Rumelhart (1977), o el modelo en cascada de McClelland (1976) del que derivaría el conocido modelo de reconocimiento de palabras de McClelland y Rumelhart (1981; Rumelhart y McClelland, 1982; hoy Seidenberg y McClelland, 1989) deben ser citados, entre otros muchos, como claros antecesores del conexionismo actual.

Sin embargo, el interés masivo por tal enfoque vendrá tras la aparición del libro de Hinton y Anderson (1981) «Parallel models of associative memory». La crítica de Minsky y Papert será finalmente resuelta en el trabajo de Ackley, Hinton y Sejnowsky (1985). Un año después Rumelhart, Hinton y Williams (1986) desarrollaron el algoritmo de «propagación hacia atrás» o «regla generalizada delta» que resolvió definitivamente y de modo óptimo el tratamiento de las unidades ocultas. Por fin en 1986 aparecen los dos volúmenes del libro de McClelland y Rumelhart que supondrán la popularización definitiva del nuevo enfoque, considerado por autores como Palmer (1987) o Schneider (1987) como un «nuevo paradigma» en Psicología.

Coincidiendo temporalmente con la publicación de los dos volúmenes del libro del grupo PDP (McClelland y Rumelhart, 1986; Rumelhart y McClelland, 1986; así como su manual de uso de los programas de simulación; McClelland y Rumelhart, 1988) el conexionismo se ha puesto de moda entre la comunidad de psicólogos actuales. Una mera revisión de las revistas especializadas sobre computación (véase, p. e., Zeidenberg, 1987) o de las mesas específicas organizadas en prestigiosos Simposiums como los de la «Cognitive Science Society» o «Psychonomic Society» dan fe del boom bibliográfico que el conexionismo despierta entre los estudiosos de la genéricamente llamada «Ciencia Cognitiva».

En este sentido desde mediados de la década actual las aportaciones conexionistas crecen día a día, tanto en el campo de la Psicología como en el de la computación e Inteligencia Artificial (IA). Aportaciones conexionistas en el campo psicológico abarcan áreas tan variadas como producción del lenguaje (Dell, 1988), percepción del lenguaje (McClelland y Elman, 1986a y b; Seidenberg y McClelland, 1989), percepción del tiempo (Church y Broadbent, en prensa), aprendizaje asociativo (Gluck y Thompson, 1987), simulación de la conducta motora (Bullock y Grossberg, 1988; Rumelhart y Norman, 1982), adquisición de destrezas (Schneider y Oliver, 1988), modelos de atención (Schneider y Shedden, 1989), modelos de memoria (McClelland y Rumelhart, 1985, 1986a) y simulación de procesos amnésicos (McClelland y Rumelhart, 1986b), percepción de objetos (Grossberg y Mingolla, 1985, 1986), o pensamiento (Graham y Schneider, 1988; Rumelhart, Smolensky, McClelland y Hinton, 1986). En un campo más puramente informático podríamos citar aportaciones como sistemas de procesamiento del lenguaje natural (Sejnowsky y Rosenberg, 1987; Waltz y Pollack, 1985), sistemas expertos (Mozer, 1987), interfases entre sistemas informáticos (Jones, 1987), y optimización de problemas (Hopfield y Tank, 1985), entre otros muchos.

Sin embargo, al margen de sus antecedentes históricos, cabría preguntarse sobre ¿qué aporta de nuevo el conexionismo en relación a las formas anteriores de abordar la implementación de aspectos cognitivos, como para que haya adquirido una importancia explosiva en un período muy breve de tiempo? Centraremos la respuesta a este interrogante desde una triple vertiente:

a) De un lado por lo que respecta a la *representación del conocimiento*, frente al carácter simbólico de la información propia de los programas tra-

dicionales de la IA y Ciencia Cognitiva, donde cada unidad de procesamiento suele representar algo en concreto (conceptos, palabras, fonemas, imágenes, etc.; o «representación local o simbólica del conocimiento», Estes, 1988; Smolensky, 1988; con aportaciones clásicas como las de Anderson, 1976; Fodor, 1975, Newell, 1980 o Pylyshyn, 1984), el enfoque conexionista opta por un *carácter más atomista* de la representación, donde las unidades no suelen representar nada en concreto, sino que el significado surge de la interacción de conjuntos muy amplios de unidades. Es decir, muy usualmente en los modelos conexionistas el conocimiento simbólico se representa como una pauta particular de activación entre un conjunto de unidades en un determinado momento temporal: de igual modo que podemos representar un concepto determinado adscribiéndolo a una única unidad también podríamos representarlo aquí como un determinado valor de activación en un conjunto más o menos amplio de unidades. De ahí el calificativo con el que también se conoce a estos modelos como de *modelos de memoria distribuida en paralelo*. Smolensky (1988) habla de ellos como modelos de representación subsimbólica. Toda conducta, por compleja que sea, sería reductible en último término a tal juego interactivo de miles de elementos muy simples.

Pese a que la distinción entre representación local vs. distribuida parece clara, en último término, y como señalan Oden (1988) o Hinton (en prensa) la distributividad es un problema de perspectiva y grado. A pesar de que no es requisito indispensable el admitir una representación distribuida del conocimiento para poder hablar de conexionismo, sí es cierto que una amplia mayoría de modelos conexionistas optan por la alternativa distribuida. Ello por varias razones (Hinton, McClelland y Rumelhart, 1986):

En primer lugar podemos representar gran cantidad de información en un espacio menor de memoria del que necesitaban los modelos localistas, donde se requiere búsquedas masivas, y por ende, lentas hasta encontrar el ítem que se adecúa a determinada descripción. Aquí diferentes ítems corresponden a diferentes pautas de activación sobre un mismo conjunto de unidades. Es pues posible representar nuevos ítems sin necesidad de incrementar el número de unidades.

De lo anterior se deriva otra ventaja todavía más importante cual es dotar al modelo con capacidad de generalización del conocimiento, un problema de importancia primaria para muchos teóricos en el campo de la psicología. Con ello, y esta es la ventaja realmente importante, se comienza a poder pensar en simular el carácter reconstructivo y no puramente mecánico de la memoria humana.

Por último hay que decir asimismo que tal carácter atomista no es planteado por sus defensores como un enfoque incompatible y alternativo a los otros modos de entender la ciencia cognitiva (modelos que, desde este punto de vista, se encargarían de las «macroestructuras» de la cognición) sino que el enfoque conexionista, por su parte, se centraría más en el nivel de análisis de la interacción de las estructuras mínimas, de inspiración pseudo-neuronal, que configuran la cognición, es decir de la «*microestructura*» de la cognición. Pero, como señalan Rumelhart, Hinton y McClelland (1986), muchos de los constructos situados a un nivel macro-descriptivo como esquemas, imágenes, producciones, etc. podrían ser en el futuro entendidos como derivados de las interacciones de miles de tales microestructuras.

En resumen, representaciones simbólicas y localistas frente a represen-

taciones subsimbólicas y distribuidas. Aquí aparece pues la primera aportación nueva del enfoque del que hablamos en relación a los enfoques clásicos de simulación de la conducta inteligente.

b) Más evidente aún son las diferencias entre los modelos clásicos y los modelos conexionistas en torno a su *concepción del proceso de aprendizaje*. El enfoque conexionista ha surgido como una alternativa clara a las limitaciones de los programas de simulación cognitiva hasta ahora más popularizados, y en especial aquellos basados en los llamados sistemas de producciones (p. e., J. R. Anderson, 1983). Si aprender para una red era antes mejorar e incluir nuevas producciones que optimizaran los resultados (con las limitaciones que imponían hechos como el aumento masivo de producciones, el que pudieran llegar a entrar en contradicción unas reglas con otras, tener que establecer jerarquías de aplicación entre ellas, etc.) ahora será simplemente modificar las conexiones existentes entre las unidades. Este hincapié en el carácter continuamente ajustable de las conexiones define un rasgo idiosincrático del enfoque conexionista. Para ello no se propugnará un almacén (procedimental) ajeno a la red, sino que ésta contará con los recursos inherentes necesarios para posibilitar de forma automática la modificación de las mismas a través de la experiencia y su ajuste para posibilitar alguna computación específica. En el conexionismo las redes «se programan a sí mismas» (Smolenky, 1988), dado que la red cuenta con los algoritmos adecuados para ello (p. e., la regla de Hebb, regla delta, regla generalizada delta; véase más abajo). Debido a tal carácter ajustable de las conexiones las redes no son concebidas como algo estático, sino entidades en un continuo cambio tendente a optimizar el aprendizaje. El desarrollo de algoritmos de aprendizaje así entendidos está a la base de todo el desarrollo teórico del conexionismo (Hinton, en prensa).

Paralelismo, atomismo y modificación continua de las conexiones a través del aprendizaje son pues los tres pilares básicos sobre los que descansa el edificio conexionista. Intentar, sin embargo, dar más rasgos definitorios comunes a tales modelos supone caer en la descripción de modelos concretos, por lo que no ahondaremos más sobre tal punto.

c) Es sin embargo la *inspiración neuro-fisiológica* en la que se basan los desarrollos teóricos conexionistas el hecho más novedoso en relación a los modelos anteriores de simulación. Ya hemos citado la evidente similitud formal de los modelos conexionistas con la anatomía externa y funcionamiento general del córtex cerebral. El cerebro consiste en un gran número de elementos muy simples altamente interconectados entre los que transitan mensajes excitatorios e inhibitorios y que adaptan sus conexiones en base a tales mensajes. Sin embargo, y pese a que se admite explícitamente el influjo de las aportaciones neurofisiológicas de la anatomía cerebral, hay que decir que tales modelos no son, ni pretenden ser, modelos de simulación del funcionamiento del córtex, a excepción de algunos modelos específicamente desarrollados para ello (p. e., McNaughton y Morris, 1987). Hoy por hoy, y como vamos a ver, la evidencia neurofisiológica no pasa de ser un símil más o menos inspirativo a la hora de formular los programas, dado que las diferencias entre uno y otro sistema son evidentes (Smolensky, 1988):

- Los ordenadores procesan la información de un modo mucho más rápido a como lo hacen las neuronas. Mientras que la unidad de medida del tiempo en aquéllos se refleja en nanosegundos, las neuronas

- parecen trabajar en milisegundos o incluso centésimas de segundo. De este modo se podría afirmar que los ordenadores son del orden de 10^6 veces más rápidos procesando la información que el sistema neuronal.
- El número de neuronas del córtex es de 10^{10} a 10^{11} , número hoy por hoy inabordable para una máquina. Sin embargo autores como Schneider (1987) son optimistas al respecto al señalar que en el plazo de 5-10 años podremos construir sistemas que emulen redes de millones de nodos interconectados por millones de asociaciones.
 - Cada neurona se asume que tiene entre mil a cien mil sinapsis con otras neuronas, mientras que en las redes artificiales actuales el número de conexiones se mide tan sólo por decenas.
 - Las conexiones del córtex cerebral son muy diferentes: algunas son muy cortas (las más), mientras que otras (las menos) suelen ser muy largas. Además suelen ser simétricas. Tales aspectos no son tenidos en cuenta, hoy por hoy, en los modelos de simulación conexionistas.
 - Como tampoco tienen en cuenta los modelos conexionistas aspectos tan importantes en el funcionamiento del córtex como pueden ser los efectos de agentes neuro-químicos, o la continuada mortandad neuronal que desde muy pronto en él acaece, etc.

Pese a todas estas diferencias (muchas de ellas probablemente debidas sólo al estado inicial de desarrollo en que se encuentran los sistemas informáticos actuales, es decir, a limitaciones tecnológicas), bien es verdad que el sistema neurofisiológico y modelos conexionistas de simulación sí coinciden en una serie de aspectos importantes, además de los reseñados. Por ejemplo:

- En ambos sistemas se asume que el funcionamiento global depende de la interactuación de una enorme cantidad de unidades sencillas cuyo estado continuamente cambia en el tiempo.
- Unos y otros asumen que ninguna unidad en concreto es imprescindible para el funcionamiento global del sistema, tal y como apuntaron Lashley y Luria. Pese a que se puede admitir que regiones específicas del córtex son las encargadas de llevar a cabo funciones concretas (particularmente a niveles bajos de procesamiento), se podría asumir que el funcionamiento del cerebro descansa en la interactuación en paralelo de millones de neuronas, sin que de ninguna en particular sea la responsable del rendimiento del grupo.
- En uno y otro sistemas parece no haber una región encargada de la supervisión del funcionamiento adecuado del sistema en su totalidad. Más bien parece que el producto final de la acción descansa en el funcionamiento de un número altísimo de unidades simples que llevan a cabo tareas muy sencillas.

Por vez primera, pues, se intenta conectar los niveles de análisis computacional, cognitivo y fisiológico en un sistema sencillo común (Schneider, 1987), siendo de esperar que, en un futuro no muy lejano, puedan terminar por hablar en un lenguaje común. Nunca hasta ahora informática, psicología y fisiología habían estado trabajando en áreas tan próximas. Sin embargo, y como señalan los principales responsables de tal enfoque, se hace necesaria una perseveración continua y sostenida en tal línea de acercamiento.

Planteados los postulados teóricos básicos que configuran lo que hoy, grosso modo, entendemos por conexionismo entraremos ahora en una consideración formal más detallada de los componentes de una arquitectura conexionista.

2. ELEMENTOS DE UNA ARQUITECTURA CONEXIONISTA

El primer paso en la elaboración de un modelo de simulación consiste en la elección de una determinada arquitectura que disponga los elementos de alguna forma específica en relación con el fin que se pretenda conseguir, junto con la elección de los mecanismos computacionales que van a dar lugar al procesamiento deseado. Ambos aspectos, estructura y mecanismos de procesamiento se concretan en el patrón de unidades e interconexiones entre las mismas, estados de activación, y formas de transformación de las interconexiones entre las unidades. A continuación procedemos a analizar cada uno de estos aspectos.

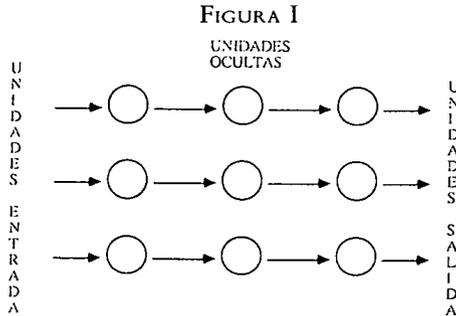
2.1. La estructura de una red conexionista

Un modelo conexionista está compuesto por una serie de unidades computacionales con distintas funciones y en una distinta configuración asociativa. Todas las unidades vienen caracterizadas (en cada uno de los ciclos en que se descompone la variable «tiempo») a través de un nivel de activación determinado, de igual modo que cada conexión viene definida por su peso correspondiente. Los niveles de activación permitidos a las unidades (discretos o continuos, con límites o sin ellos,...), y a los pesos de sus asociaciones (discretos o continuos, con límites o sin ellos, sólo positivos—en cuyo caso hablaremos de conexiones excitadoras— o también negativos—en cuyo caso hablaremos de conexiones inhibitorias—, ...), son el primer criterio diferenciador de los distintos modelos conexionistas. Todo el conocimiento suele representarse en un sistema conexionista en función de entre qué unidades se dan conexiones y el valor de las mismas (Hinton, en prensa) de tal modo que, como vamos a ver, aprender no será sino el proceso por el que creamos nuevas asociaciones o modificamos las ya existentes.

Desde un punto de vista formal hay varios tipos de caracterizar estructuralmente una red conexionista. Hay que decir, ya de entrada, que por el momento (en base probablemente al estado inicial de desarrollo en que se encuentra este campo) no existe ningún algoritmo o procedimiento que dicte qué configuración elegir ante un problema dado. Es el programador el que bajo criterios puramente subjetivos debe elegir entre uno de los siguientes tipos de arquitecturas:

2.1.1. Las más popularizadas son los llamados *asociadores de patrones* (pattern associator; McClelland, Rumelhart y Hinton, 1986). En ellos es útil distinguir tres tipos de unidades: A) *Unidades de input o de entrada* o unidades con capacidad de ser activadas desde el exterior de la red. A las activaciones que reciben tales unidades desde fuera del sistema se les denomina «patrones de entrada». B) *Unidades de output o de salida* o unidades con

capacidad de mostrar la respuesta del sistema a cierto input. A las activaciones que es de esperar tengan tales unidades como respuesta a las activaciones provocadas por ciertos patrones de entrada se les denomina «patrones de salida». C) *Unidades ocultas* (opcionales), o unidades que ni pueden ser activadas desde el exterior ni pueden dar lugar por sí solas un output, sino que sólo sirven de intermediarias entre los tránsitos de activación que acaecen desde las unidades de entrada hacia las de salida. A los asociadores de patrones en cuya configuración entre al menos una capa de unidades ocultas se los conoce como sistemas multicapa (véase Figura 1).



*Ejemplo simbólico de la arquitectura de un asociador de patrones multicapa.
En realidad todas las unidades se encuentran asociadas con todas.
Symbolic example of a pattern associator architecture with hidden units.*

Si a un asociador de patrones le presentamos un conjunto de patrones de entrada y el conjunto de patrones de salida que queremos asociar a aquéllos, y si la red cuenta con los algoritmos de modificación de las conexiones adecuados (véase más abajo), debería llegar un momento en que la red fuese capaz de producir por sí sola el patrón de salida asociado a determinado patrón de entrada. Este es el objetivo básico de todo asociador de patrones: asociar en una única matriz de conexiones un conjunto de patrones de entrada y un conjunto de patrones de salida. La relativa sencillez de estos planteamientos se ve empañada por las limitaciones que plantean los algoritmos usualmente utilizados para modificar las conexiones de la red (reglas de Hebb y delta, de las que luego hablaremos). Como mostraron Minsky y Papert (1969) en un sistema sin unidades ocultas no siempre es posible asociar de un modo óptimo un patrón de entrada a un patrón de salida. En base a tales limitaciones se han desarrollado los sistemas multicapa (Figura 1). El conexionismo ha puesto considerable énfasis en el «problema de las unidades ocultas» (Hinton y Sejnowsky, 1986). Algoritmos de aprendizaje desarrollados recientemente (y en especial la llamada regla delta generalizada o regla de propagación hacia atrás) permiten computar los niveles de activación de todas las unidades y el valor de las conexiones que las asocian a lo largo de los distintos ciclos, de tal modo que con ellos es prácticamente abordable cualquier tipo de problema asociativo.

2.1.2. Otro tipo de arquitectura conexionista muy popularizada es aquella en la que todas las unidades quedan asociadas con todas (a veces incluso consigo mismas), no cabiendo distinguir entre unidades de entrada y de salida, de tal modo que los inputs hacia el sistema se proyectan sobre ciertas

unidades y el output es medido en relación a los niveles de activación que dichas unidades han alcanzado. Se podría decir pues que todas las unidades de la red hacen las veces de unidades de entrada y unidades de salida. En base a los objetivos que se persigue con este tipo de arquitectura cabría hablar de:

2.1.2.1. *Autoasociadores* cuya finalidad es la misma que la de los asociadores de patrones, es decir, aprender a asociar patrones de entrada a específicos patrones de salida a través de los algoritmos de modificación de las conexiones que luego veremos. Cabe hablar a su vez de distintos tipos de autoasociadores (véase p. e., McClelland y Rumelhart, 1988) como el *lineal* (donde la activación de una unidad determinada es la suma de las activaciones de las unidades vecinas ponderadas por los pesos de las conexiones que las unen), el *modelo BSB* (de Brain State in the Box; véase J. A. Anderson et al., 1977; en el que se constriñe el máximo y el mínimo nivel de activación que las unidades pueden alcanzar), el *modelo DMA* (de Distributed Memory and Amnesia; véase McClelland y Rumelhart, 1985, 1986a, 1986b, 1988, cap. 2; aplicado a la simulación del funcionamiento de la memoria humana), etc.

2.1.2.2. Cuando el objetivo que se persigue con una red en la que todas las unidades quedan asociadas con todas es la detección de regularidades entre un conjunto de patrones de entrada y/o satisfacer de modo óptimo determinados requerimientos de búsqueda (constraint satisfaction models) cabe hablar entre otros de las *máquinas de Boltzman* (Boltzman machine; véase Hinton y Sejnowsky, 1986), *modelo de esquema* (Schema model; véase Rumelhart, Smolensky, McClelland y Hinton, 1986), etc.

2.1.3. Otras arquitecturas jerarquizan distintas capas de unidades (interactuantes entre sí) adscribiendo a cada una de ellas un significado representacional distinto (p. e., una capa de unidades para representar rasgos físicos del lenguaje escrito, otra para representar un lexicón, otras para representar una red conceptual, etc.). De este modo encontramos entre otras la *teoría armónica* (Harmony theory; Smolensky, 1986, cuyo fin es también maximizar la satisfacción de una demanda en la búsqueda), el modelo de *aprendizaje competitivo* (competitive learning o modelo jerárquico para la detección de regularidades entre patrones; véase Rumelhart y Zipser, 1986) o el *modelo IA* (de interactive activation, aplicado a un modelo de percepción y producción del lenguaje; véase McClelland y Rumelhart, 1981; Rumelhart y McClelland, 1982; Seidenberg y McClelland, 1989).

Por razones obvias se sale fuera del propósito de este trabajo una consideración exhaustiva de tales aportaciones. Debemos aclarar asimismo que la taxonomía de modelos propuesta no pretende ni ser exhaustiva ni exclusiva en el sentido de que como Rumelhart, Hinton y McClelland (1986) señalan muy comúnmente aspectos de unas y otras arquitecturas suelen mezclarse en una simulación concreta. En base a la mayor importancia relativa que las arquitecturas de asociador de patrones reciben (véase McClelland, Rumelhart y Hinton, 1986) centraremos desde ahora nuestros comentarios en torno a ellos.

2.2. Mecanismos de procesamiento

En este apartado abordaremos el tema de las funciones por las cuales las unidades van variando su nivel de activación y las conexiones van modificando su peso a lo largo de los distintos ciclos iterativos de cálculo.

Pese a que como veremos también aquí aparecen diferencias notables entre los distintos modelos conexionistas, todos ellos asumen que la activación que una unidad concreta j en un momento dado tiene depende de: su nivel de activación previo, la activación de las n unidades con las que está relacionada y de las pautas de conectividad que se dan entre ellas. Distintas funciones son aplicadas para dar fe de tal proceso si bien todas ellas derivan de una idea similar, cual es que el input que recibe una unidad j (net_j) no es sino la suma de la activación de las n unidades asociadas a ella ponderada por el valor de las conexiones que las unen.

$$net_j = \sum_{i=1}^n w_{ij} a_i$$

Tal net_j usualmente es añadido al nivel de activación que tal unidad j tenía en un momento temporal anterior para computar su nuevo nivel de activación (función lineal), si bien otras veces se puede emplear una función discreta (donde sólo se activaría dicha unidad cuando el net que recibiera sobrepasase cierto criterio), sigmoide continua, etc. El problema de utilizar la función lineal estriba en que da lugar a crecimientos explosivos de la activación de todas las unidades debido a las iteraciones continuas (McClelland, 1988) a menos que las conexiones sean constreñidas de algún modo (p. e., incluyendo una tasa de aprendizaje-decaimiento en las mismas; o a través de la propuesta del modelo Brain-State-in-the-Box, comentando).

Al igual que cuando hablamos de cómo poder transformar los inputs de las unidades en activaciones, cabe hablar de las distintas funciones para computar la activación que una unidad enviará a sus vecinas a partir de su estado actual. Aquí también las tres funciones de activación más comúnmente utilizadas son (Oden, 1988): lineal, de umbral y logística semilineal. Debemos decir que pese a que la gran mayoría de modelos conexionistas aplican estas funciones de activación de forma sincrónica (en cada ciclo temporal todos los niveles de activación son readaptados en función de los inputs que reciben) otros modelos, como el modelo de esquema, las aplican de forma asincrónica (en el sentido de que sólo son readaptados los niveles de activación de algunas unidades tomadas al azar).

Sin embargo la aplicación y desarrollo de las reglas de *modificación de las conexiones* son, como ya se comentó arriba, el punto crucial de la investigación en los modelos conexionistas. Como queda dicho, si el conocimiento se representa entre las conexiones, el aprendizaje no será sino el proceso por el que tales conexiones cambian. Aprender consiste en modificar las conexiones existentes entre un conjunto de unidades.

Una amplia variedad de «reglas de aprendizaje» han sido propuestas para lograr tal fin (regla de Hebb, regla delta, regla del aprendizaje competitivo, regla delta generalizada...; las dos primeras características del aprendizaje en asociadores de patrones simples, las restantes relativas a redes donde intervienen unidades ocultas; véase Fogelman y Mejía, 1988 o Hinton, en prensa, para una revisión exhaustiva sobre el tema). Generalmente todas ellas asumen que

la modificación de una conexión depende del producto de dos términos (McClelland, 1988): un término pre-sináptico asociado con la activación de la unidad aferente y un término post-sináptico asociado con la activación de la unidad eferente. El problema radica pues en hallar qué matriz(es) de conexiones permite(n) asociar de un modo óptimo unos valores de activación del conjunto de unidades de entrada (o patrones de entrada) con unos valores de activación del conjunto de unidades de salida (o patrones de salida), de tal modo que en el futuro siempre que presentemos uno de aquéllos el sistema permita recuperar del modo mejor posible el patrón de salida que se le asoció. Por su relevancia bibliográfica incidiremos especialmente en el comentario de las reglas Hebb, delta y delta generalizada.

La *regla de Hebb* (también llamada «regla del producto») establece que el incremento que experimenta el peso de la conexión que asocia al unidad i con la unidad j (ΔW_{ij}) es directamente proporcional al producto de sus niveles de activación ante cada ciclo temporal:

$$\Delta W_{ij} = a_i a_j e$$

siendo e un parámetro constante de aprendizaje-decaimiento.

Se puede demostrar que distintos patrones de activación de entrada podrán dar lugar a distintos valores de conexión con respecto a las mismas unidades de salida si aquéllos no son ortogonales entre sí (Hinton, en prensa; Pitarque, Algarabel, Dasí y Ruiz, 1989). En otras palabras, cuando el requisito de ortogonalidad no es satisfecho por los vectores de entrada, la regla de Hebb no da lugar a soluciones únicas, en el sentido de que un mismo par de vectores de salida y entrada pueden quedar asociados por medio de distintos valores de conexiones. Una forma de solventar este problema es apelar al uso de vectores de entrada con valores aleatorios y con un número muy grande de dimensiones. Pero ello no siempre es posible o deseable.

Una forma de salvar estas limitaciones puede ser apelar al uso de la *regla delta* (también llamada «regla del menor error cuadrático») basada en el cómputo y reducción constante de la diferencia cuadrática existente entre los vectores de salida que se desea alcanzar y los que son producidos en cada ciclo iterativo por la red. Intuitivamente hablando el objetivo de la regla delta es minimizar de forma gradual la discrepancia entre los patrones de salida que produce la red en cada ciclo de iteración con el patrón de salida que se desea conseguir; cuando tal discrepancia es nula la regla deja de aplicarse. En otras palabras, cuando a lo largo de los distintos ciclos iterativos de cómputo, la suma de las diferencias cuadráticas correspondientes a todas las unidades de salida con respecto a los valores que han de alcanzar es 0, el proceso finalizada, dándose por acabado el ajuste de los pesos. Se dice entonces que la red está equilibrada. Obsérvese cómo mientras la regla de Hebb está concebida para hacer los ajustes en un solo paso la regla delta sigue un procedimiento iterativo de ajuste en términos de un gradiente decreciente.

La regla delta ajusta las conexiones de acuerdo a la siguiente expresión:

$$\Delta w_{ij} = e a_i e_j$$

donde a_i es la activación de la unidad de entrada en un ensayo concreto,

e_j es el error para la unidad de salida j en ese ensayo que viene dado por el cómputo previo de

$$e_j = (t_j - a_j)$$

o diferencia entre la activación que sería de esperar para dicha unidad caso de darse un aprendizaje óptimo (t_j) y la que en este momento temporal realmente tiene dicha unidad (a_j).

La *regla generalizada delta* (o regla de propagación hacia atrás) fue desarrollada por Ackley, Hinton y Sejnowsky (1985) y Rumelhart, Hinton y Williams (1986) para poder resolver las limitaciones de los asociadores de patrones sin unidades ocultas (p. e., la resolución de ciertas relaciones lógicas que las anteriores reglas no pueden resolver). Apelando al uso de unidades ocultas es posible hallar siempre los valores de conexiones que asocien de un modo óptimo cualquier patrón de entrada con cualquier patrón de salida, independientemente de las características de éstos.

De forma intuitiva el significado de esta regla es similar al de la regla delta, sólo que aquí se trabaja con redes con unidades ocultas. El procedimiento general de la regla generalizada delta es el siguiente: Presentamos al sistema un conjunto de pares de patrones de entrada y salida; dejamos que en cada caso la presentación de un patrón de entrada dé lugar a su propio vector de salida y luego comparamos este resultado con el patrón de salida que en realidad se le quiere asociar. Si ambos outputs no difieren entonces no habrá aprendizaje. En otro caso los pesos son modificados para reducir tales diferencias en forma de un gradiente descendente a lo largo de los ciclos iterativos de cálculo. Si trabajásemos con sistemas sin unidades ocultas estaríamos hablando aquí de la regla delta ya conocida. Sin embargo estamos hablando de sistemas con al menos una capa de unidades ocultas.

El problema se resuelve según Rumelhart et al (1986) del siguiente modo: en primer lugar presentaremos a la red el input de entrada y computaremos los outputs (a_{pi} para cada unidad de salida i en cada ciclo de procesamiento p) a través de la función de activación pertinente. La regla exige que exista la derivada de la función de activación que utilicemos. Se necesita pues apelar al uso de una función continua y no lineal. Para ello McClelland y Rumelhart (1988) recomiendan utilizar la función de activación semilineal sigmoide del tipo

$$a_{pi} = \frac{1}{1 + e^{-net_{pi}}}$$

Luego compararemos tales valores a_{pi} con los valores que desearíamos alcanzar (t_{pi}) si la asociación entre el patrón de entrada y de salida fuese perfecta resultando para cada unidad de salida un valor de error δ_{pi} . Tal valor viene dado por:

$$\delta_{pi} = (t_{pi} - a_{pi}) f'(net_{pi})$$

donde net_{pi} es el input que cada unidad i recibe de las otras.

En el caso de la función de activación propuesta por McClelland y Rumelhart (1988) podemos derivar y sustituir arriba

$$\partial_{pi} = (t_{pi} - a_{pi}) a_{pi} (1 - a_{pi})$$

El segundo paso en la aplicación de la regla de la que hablamos envuelve la computación «hacia atrás» de ∂_{pi} no ya sólo para las unidades de salida, sino para todas las unidades ocultas, en términos de los ∂_{pk} de las unidades de salida con las que cada unidad oculta se conecta y de sus conexiones. De este modo el error para la unidad oculta i en el ciclo p será:

$$\partial_{pi} = f'(\text{net}_{pi}) \sum_{k=1}^n \partial_{pk} w_{ki}$$

derivando la función de activación comentada y sustituyendo

$$\partial_{pi} = a_{pi} (1 - a_{pi}) \sum_{k=1}^n \partial_{pk} w_{jk}$$

Una vez que estas dos fases han sido completadas se computa para cada conexión el producto del término ∂_{pi} por la activación de la unidad que asocia, en modo similar a como acontecía en la aplicación de la regla delta:

$$\Delta W_{pji} = e \partial_{pi} a_{pj}$$

siendo e el parámetro constante de aprendizaje-desactivación.

Este algoritmo se ha aplicado a muchos problemas: reconocimiento de imágenes y palabras, identificación de señales, redes semánticas. Sin embargo, gran número de problemas quedan, hoy por hoy, sin solución como su rendimiento diferencial ante distintas arquitecturas neurales, ante los distintos parámetros de la regla o ante la matriz de pesos iniciales (Fogelman y Mejía, 1988).

3. SIMULACION DE LOS EFECTOS DE FACILITACION SEMANTICA

Objetivo

Como es bien sabido el procesamiento de un estímulo facilita la recuperación de un estímulo relacionado con él. Por ejemplo, en las llamadas tareas de decisión léxica se evidencia cómo los tiempos de reacción que invierten los sujetos en decidir si un conjunto de letras forman o no una palabra suelen ser inversamente proporcionales al grado de asociatividad que tal estímulo test mantiene con la palabra que le precedió en la lista de presentación. Tal facilitación semántica debida a la relacionalidad entre conceptos (conocida como «efecto de priming»; véase, p. e., Meyer y Schavaneveldt, 1971) suele ser tomada como evidencia a favor y con carácter de exclusividad de los modelos que propugnan una consideración localista y simbólica de la memoria (véase, p. e., Anderson, 1983; Collins y Loftus, 1975). En este sentido, y como señala Neely (1990), modelos tan prototípicamente conexionistas como el de reconocimiento de palabras de Seidenberg y McClelland (1989) tiene serias dificultades para simular tal efecto de priming semántico (véase también Besner et al., 1989, para evidencia experimental en contra del modelo de

Seidenberg y McClelland), lo que, junto con otros resultados experimentales que huelga comentar aquí, es interpretado por dicho autor como sería evidencia en contra de tales modelos PDP (véase también Massaro, 1988, 1989; o Ratcliff, 1990 como representantes de una postura abiertamente anti-conexionista).

El objetivo de la presente simulación es intentar demostrar cómo desde una sencilla arquitectura conexionista (en este caso, y a diferencia del modelo de Seidenberg y McClelland, un asociador de patrones) es posible llegar así mismo a simular tales resultados de facilitación, lo que evidenciaría que los llamados modelos activacionales clásicos no tienen competencia exclusiva en la explicación de los mismos, en contra de lo que afirma, entre otros, Neely (1990). Este dato, junto a hechos como el que tales modelos simbólicos no tienen una evidencia neurofisiológica seria que los avale (p. e., nadie ha localizado, ni lo hará, donde se representa corticalmente el llamado sistema de producciones), o que necesiten postular continuamente y «a posteriori» almacenes y estructuras mnésicas de dudoso origen para dar fe de la distinta evidencia experimental que va surgiendo (p. e., términos como memoria a corto plazo/memoria a largo plazo, niveles de procesamiento, propagación de la activación, memoria semántica/episódica, almacén declarativo/almacén procedimental han dado lugar a mucha bibliografía en las pasadas décadas pero a resultados predictivos poco esperanzadores) dejaría a tales modelos en clara inferioridad, desde nuestro punto de vista, con respecto a la potencia explicativa de los modelos PDP.

Procedimiento

Como en casi todas las arquitecturas conexionistas consideraremos un concepto cómo una particular pauta de activación distribuida entre un conjunto de unidades en un momento temporal dado (y representada a través de un vector de activación). El problema que se nos planteaba en esta simulación era cómo operacionalizar el grado de asociación semántica entre los conceptos. Se podría pensar que un concepto sería igual a otro cuando las pautas de activación entre el conjunto de unidades que los representaran fueran similares. De igual modo se podría pensar que a medida que dichas pautas fueran difiriendo, en ese grado disminuiría la relación existente entre ambos conceptos. De este modo tomaremos el grado de distorsión incorporada a una particular pauta de activación como indicadora del alejamiento asociativo (o «distancia semántica» en términos de Collins y Loftus) existente entre el concepto representado por la pauta original y el representado por la pauta distorsionada de aquélla.

Para simular el paradigma experimental de priming decidimos elaborar un programa que creaba un asociador de patrones de 15 unidades de entrada asociadas a 15 unidades de salida (sin unidades ocultas) asociadas a través de una matriz de conexiones. Tanto el nivel de activación inicial de las unidades como los valores de los pesos que las asociaban tomaron valores aleatorios comprendidos entre $+/-1$. Una vez definidas la arquitectura inicial de la red sometimos al aprendizaje de 9 pares de vectores entrada y 9 vectores de salida (con valores aleatorios comprendidos entre $+/-1$), presentados en orden secuencial y aleatorio a lo largo de 15 ciclos iterativos

(donde cada ciclo estaría por cada uno de los valores discretos en que se operacionaliza la variable «tiempo» en estos modelos), y que guardaban entre sí la siguiente relación:

- En tres de los nueve ensayos se repetía el mismo par de vectores entrada-salida. En el paradigma experimental de decisión léxica tal condición haría las veces de condiciones de repetición señal-test, por lo que llamaremos en nuestra simulación a tal condición de repetición (REP).
- En otros tres ensayos se presentaban vectores de entrada distintos (aleatorios, y distintos en cada ciclo) asociados a un mismo vector de salida. En las tareas de decisión léxica tal caso equivaldría a las condiciones de no relacionalidad entre estímulos. Aquí la llamaremos de no relacionalidad (NO REL).
- En los otros tres ensayos los patrones de entrada eran distorsiones del patrón de entrada que se presentó en la condición REP, asociados al mismo vector de salida que el utilizado en aquel caso. Es una forma de evidenciar una cierta relacionalidad entre conceptos. Por ello llamaremos a tal condición condición de relacionalidad señal-test (REL). Las distorsiones de los vectores de entrada en relación al vector de entrada fuente se lograron modificando los valores de activación iniciales de todas las unidades de entrada en relación a incrementos o decrementos aleatorios comprendidos entre $+/- .20$ (REL .20), $+/- .40$, ó $+/- .60$ (REL .60).

La hipótesis que se plantea es pues poner a prueba si el programa es capaz o no de simular los resultados experimentales conocidos de mejor ejecución en los casos en que la relación señal-test es más estrecha (es decir, una facilitación establecida en este orden $REP > REL.2 > REL.4 > REL.6 > NO REL$) tal y como acaece en la literatura experimental en sujetos humanos.

La simulación se llevó a cabo a través de un programa escrito en QUICK-BASIC de Microsoft (versión binaria 2.00) e implementado en un ordenador Macintosh II de Apple. Se utilizaron funciones de estado y propagación de la activación lineales (es decir, donde la activación de una unidad en un ciclo dado es linealmente proporcional a la activación de las unidades con las que está asociada ponderada por los pesos de las asociaciones que las unen) y la regla de aprendizaje delta, arriba comentada, para la modificación de las conexiones. Para evitar crecimientos explosivos de la activación el incremento de los pesos de las conexiones determinado por la aplicación de tal algoritmo de aprendizaje era ponderado por una constante (igual a 1 dividido entre el número de unidades), mientras que el incremento que experimentaba cada valor de activación de cada unidad era asimismo ponderado por otro parámetro ($= .15$).

La variable dependiente fue la suma de cuadrados del patrón (PSS) o suma de las diferencias cuadráticas existentes patrón a patrón entre las actividades de las unidades de salida y el patrón de salida que realmente se quería asociar a cada vector de entrada (véase McClelland y Rumelhart, 1988). En este contexto una $PSS=0$ indicaría un grado de aprendizaje perfecto, mientras que a medida que tal variable se alejase de tal valor evidenciaría un grado de aprendizaje más y más imperfecto. Por último, el programa se repetía el

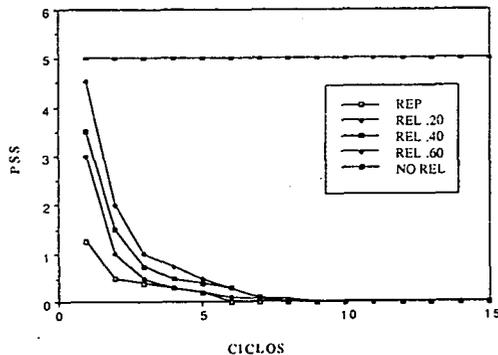
proceso anterior 15 veces, midiéndose en cada ciclo y para cada condición experimental la variable dependiente considerada.

Resultados

La Figura 2 muestra los resultados globales de nuestra simulación. En ella se puede apreciar cómo mientras la condición NO REL mantiene a lo largo de los ciclos un grado de aprendizaje nulo (en el sentido que la variable criterio permanece constante en un valor alejado de cero a lo largo de los distintos ciclos), todas las demás condiciones que engloban algún tipo de relacionalidad muestran un patrón progresivamente decreciente y tendente a 0 en su PSS. Ello evidencia ya de entrada que la red es capaz de discriminar entre patrones de entrada y salida que a lo largo de los distintos ciclos le son presentados y aprende a asociar aquellos que tienen algo en común.

Pese a que las diferencias entre las cuatro condiciones de relacionalidad no son muy marcadas (dado que son del orden de décimas en su PSS), sí tienden a ordenarse del modo previsto por nuestra hipótesis, es decir, mayor grado de asociación señal-test en la condición de repetición y valores mayores de PSS a medida que en la relación entre señal y test se dan más y más distorsiones. Es de destacar asimismo cómo tales diferencias sólo se mantienen entre los siete primeros ciclos de procesamiento, para alcanzar la red un grado de aprendizaje a partir de dicho punto. Ello es debido, lógicamente, al procedimiento de ajuste de gradiente decreciente que aplica la regla delta.

FIGURA 2



Resultados de la simulación 1 (explicación en texto).

Results of the simulator 1 (see explanation in text).

Estos resultados muestran bien claramente cómo una arquitectura que procese en paralelo representaciones distribuidas es capaz de dar fe de los mismos datos experimentales que tradicionalmente han sido tomados como evidencia en exclusiva a favor de los modelos simbólicos y localistas. Al ser los modelos conexionistas más parsimoniosos, en el sentido que tienen una clara inspiración neurofisiológica, se haría necesario un replanteamiento del poder explicativo de aquellos modelos en relación a los conexionistas.

4. VALORACION CRITICA DEL ENFOQUE CONEXIONISTA

Algunos autores hablan del nacimiento de un nuevo paradigma en Psicología (Palmer, 1987; Schneider, 1987; pero también Massaro, 1989, para un punto de vista distinto). ¿Hasta qué punto es fundado el boom que ha supuesto recientemente el conexionismo? Hay que reconocer que es bien cierto que el conexionismo ha incorporado un nuevo vocabulario y un nuevo punto de vista en la consideración de los fenómenos cognitivos. Conceptos como redes semánticas, esquemas o procesos automáticos-controlados son ahora explicados en términos de redes neurales, vectores de activación, matrices de conexiones y representaciones distribuidas. Su predictibilidad empírica (como, por ejemplo, muestra la simulación conducida arriba) y su aplicabilidad tecnológica quedan fuera de toda duda (de hecho se están desarrollando sistemas de detección de explosivos, misiles de la llamada «guerra de las galaxias», etc., a través de postulados conexionistas), lo que es muy de agradecer en disciplinas tan poco predictivas como la Psicología. Pero es sin duda el enfoque integrado de distintas disciplinas a partir del símil neurofisiológico (del que hablamos arriba) la aportación, a nuestro entender, más importante del conexionismo a las ciencias cognitivas: por vez primera se han puesto las bases para que los estudios del hardware (neurofisiológicos, etc.) y del software (psicólogos, principalmente) puedan hablar un lenguaje común.

Sin embargo hay que precisar asimismo que, como los máximos defensores de este nuevo enfoque reconocen, hoy por hoy el conexionismo plantea más interrogantes que respuestas:

Por ejemplo, el impacto empírico del conexionismo hoy por hoy ha sido mínimo, es decir, el conexionismo apenas genera experimentación. Ello debido tanto en base a la inconcreción de las predicciones de tales modelos conexionistas (con asumir los principios doctrinales del paralelismo y del carácter autónomo de la modificación de las conexiones ya podemos crear cualquier tipo de modelo conexionista) como a un hecho todavía más preocupante cual es el que los programas conexionistas, al depender de tantos parámetros arbitrariamente determinados por el programador, son cuasi-irrefutables por los datos empíricos, en el sentido de que variando el valor de tales constantes siempre es posible ajustar los resultados del programa a los resultados hallados en la literatura sobre experimentación humana. Y si este recurso tampoco funciona siempre está la posibilidad de añadir nuevas capas de unidades ocultas al sistema.

Este último punto enlaza con otra crítica seria que afecta a los modelos de los que hablamos, subyacente a la lógica de la falsación de programas de la IA y su adecuación a la evidencia psicológica humana: El que los outputs de un programa se adecúen a los resultados obtenidos en la investigación básica humana no es razón para concluir que tal programa sea un modelo trasladable a la realidad en la que se fundamentan los hechos psíquicos. Es decir, tal programa puede tener una relevancia tecnológica clara pero ninguna evidencia psicológica que lo avale. Estaríamos haciendo IA pero no Psicología (véase, p. e., Massaro, 1989, para ahondar en esta idea).

Otro punto crítico para el conexionismo actual es que los principales algoritmos de aprendizaje en los que se basa suelen optimizar los resultados ante redes y problemas pequeños, pero fracasan a la hora de lograr tales asociaciones ante redes muy grandes (Schneider, 1988). Por ejemplo, Hinton

(en prensa) demuestra que si una red de 100 conexiones necesita de 10^4 ensayos de aprendizaje para lograr un aprendizaje óptimo, una red de 1000 conexiones necesitaría de 10^7 ensayos lo que a 10 segundos por ensayos llevaría casi 14 años de funcionamiento del programa. Evidentemente el desarrollo de nuevos sistemas informáticos tendría mucho que decir en la solución a tal limitación técnica.

Los abogados del enfoque tradicional de la Ciencia Cognitiva y la IA señalan que el conexionismo sólo es predictivo para la simulación de procesos de bajo nivel (p. e., análisis de la imagen, reconocimiento y asociación de patrones, etc.) pero que es impredictivo para la simulación de constructos de más alto nivel simbólico como reglas, esquemas, procesamiento secuencial, etc. (véase, p. e., Massaro, 1988, 1989). Pese a que, como hemos visto arriba, diversos autores están afrontando desde una posición conexionista la simulación de tales conceptos, los resultados parecen dar hoy la razón a sus críticos.

Por último hay, hoy por hoy, demasiada distancia entre los modelos conexionistas y evidencia neurofisiológica de las propiedades neurales del sistema nervioso. Así coincidimos con Smolensky (1988) en asumir que:

1) Del símil del sistema nervioso aprovechamos sobre todo su estructura externa, pero no tanto sus procesos dinámicos. Sin saber qué estructuras concretas gobiernan determinados procesos, y sobre todo sin conocer qué reglas los gobiernan, podemos caer en una posición comparable a alguien que intente simular el funcionamiento del sistema solar armado con muchos datos acerca de la apariencia externa de los planetas pero si el conocimiento de las leyes de gravitación universal.

2) Hay que admitir asimismo que, en el deseo de dotar a los programas de simulación de mayor poder computacional, las decisiones tomadas a la hora de implementar tales programas parecen estar más pensando en las consideraciones matemáticas o tecnológicas que en la evidencia neurofisiológica. Así, por ejemplo, el desarrollo de reglas de aprendizaje recientes como la «regla generalizada delta» o de «propagación hacia atrás» tiene un claro anhelo predictivo pero ninguna evidencia neuronal que las avale (Hinton, en prensa). De igual modo aspectos como el carácter recursivo de la activación entre unidades (o en dos direcciones) está siendo asimismo muy cuestionado (véase, p. e., Massaro, 1989).

A la vista de estas limitaciones coincidimos con autores como Norman (1986) en asumir que en un futuro cercano es más que probable que las posiciones conexionistas y posiciones clásicas de la Ciencia Cognitiva convivan en su esfuerzo de simulación de la cognición inteligente. Aquéllas, si prosperan, accederán paulatinamente a la simulación de procesos psicológicos de cada vez más alto nivel.

EXTENDED SUMMARY

Connectionism is a method of modeling cognition as the interaction of neuron-like units. In part 1 of this article we begin reviewing (a) the historical impact (from Roseblatt to nowadays), (b) theoretical impact (connectionism models have a clear neuro-physiological inspiration and they are applied today to a huge set of research fields: Psychology, IA...) and (c) general principles (distributed representations vs localist knowledge; sub-symbolic

vs symbolic processing; parallel processing vs. serial processing, automatic learning algorithms vs production systems...) of PDP view. In part 2 we review the formal assumptions of this framework; from its main architecture (pattern associators, autoassociators..., with and without hidden units) to the principal learning algorithms (Hebb rule, delta rule and generalized delta rule). One simulation (part 3) will try to test the psychological validity of PDP models in relation to the assumptions made by the symbolic and localist models of semantic memory (Collins y Loftus, 1975; Anderson, 1983 and others), using a simple pattern associator, a linear activation function, and the delta rule. We simulated the semantic priming effect found usually in human subjects manipulating the semantic relationship between input and output vectors by means of the degree of distortion incorporated to a initial vector. Results are discussed in terms of the Healy's (1990) point of view who stated that PDP models (in particular Seidenberg and McClelland, 1989 model) could not find such priming effect. According to the results, and the more adequate neurophysiological inspiration of PDP models, we conclude that such models are more predictive than symbolic and localist models, although they are in a very early state of development.

Referencias

- ACKLEY, D. H.; HINTON, G. E. y SEJNOVSKY, T. J. (1985): A learning algorithm for Boltzman machines. *Cognitive Science*, 9, 147-169.
- ANDERSON, J. A. (1973): A theory for the recognition of items from short memorized lists. *Psychological Review*, 80, 417-438.
- ANDERSON, J. A. (1973): A theory for the recognition of items from short memorized lists. *Psychological Review*, 80, 417-438.
- ANDERSON, J. A. (1977): Neural models with cognitive implications. En LaBerge y Samuels (eds.), *Basic processes in reading perception and comprehension*. Hillsdale, N.J.: LEA.
- ANDERSON, J. A., SILVERSTEIN, J. W., RITZ, S. A. y JONES, R. S. (1977): Distinctive features, categorical perception and probability learning: Some applications of a neural model. *Psychological Review*, 84, 413-451.
- ANDERSON, J. R. (1976): *Language, memory and thought*. Hillsdale, N.J.: LEA.
- ANDERSON, J. R. (1983): *The architecture of cognition*. London: Harvard Un. Press.
- ANDERSON, J. R. y BOWER, G. H. (1973): *Human Associative Memory*. Washington D.C.: Winston.
- BESNER, D., TWILLEY, L., MCCANN, R. S. y SEERGOBIN, K. (1989): On the connection between connectionism and data: Are a few words necessary? Paper presentado en *29 Meeting Annual de la Psychonomic Society*. Atlanta.
- BULLOC, A. M. y GROSSBERG, S. (1988): Neural dynamics of planned arm movements. *Psychological Review*, 95, 49-90.
- CHURCH, R. M. y BROADBENT, H. A. (en prensa): A connectionism model of timing. En Commons, Grossberg y Staddon (eds.), *Neural network models of conditioning and action*. Hillsdale, N.J.: LEA.
- COLLINS, A. M. y LOFTUS, E. F. (1975): A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- DELL, G. (1988): The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27, 124-142.
- ESTES, W. K. (1988): Toward a framework for combining connectionism and symbolic-processing models. *Journal of Memory & Language*, 27, 196-212.
- FELDMAN, J. A. (1981): A connectionism model of visual memory. En Hinton y Anderson (eds.) (1981), *Parallel models of associative memory*. LEA: NJ.
- FODOR, J. A. (1975): *The language of thought*. London: Harvard Un. Press.
- FOGELMAN, F. y MEJÍA, C. (1988): Redes conexionistas para el aprendizaje. Tutorial presentado en el *Third Symposium on Knowledge engineering*. Madrid, octubre 1988.
- GLUCK, M. y THOMPSON, R. F. (1987): Modeling the neural substrates of associative learning and memory: A computational approach. *Psychological Review*, 94, 176-191.
- GRAHAM, D. J. y SCHNEIDER, W. (1988): Sequential learning in a connectionism model of mental arithmetic. Paper presentado en *28 Meeting Annual de la Psychonomic Society*. Chicago.

- GROSSBERG, S. (1976): Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 121-134.
- GROSSBERG, S. y MINGOLLA, H. (1985): Neural dynamics of form perception: Boundary completion, illusory forms and neon color spreading. *Psychological Review*, 92, 173-211.
- GROSSBERG, S. y MINGOLLA, H. (1986): Computer simulation of neural networks for perceptual psychology. *Behavior Research Methods, Instruments & Computers* 19, 601-607.
- HEBB, D. O. (1949): *The organization of behavior*. N.Y.: Wiley.
- HINTON, G. E. (en prensa): Connectionism learning procedures. *Artificial Intelligence*.
- HINTON, G. E. y ANDERSON, J. A. (eds.) (1981): *Parallel models of associative memory*. Hillsdale, N.J.: LEA.
- HINTON, G. E. y SEJNOVSKY, T. J. (1986): Learning and relearning in Boltzman machines. En Rumelhart, McClelland and the PDP group (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1*. Mass: MIT Press.
- HINTON, G. E., MCCLELLAND, J. L. y RUMELHART, D. E. (1986): Discussion and preview. En Rumelhart, McClelland and the PDP group (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1*. Mass: MIT Press.
- HOPFIELD, J. J. y TANK, D. W. (1985): Neural computation of decisions in optimization problems. *Biological Cybernetics*, 52, 147-152.
- KOHONEN, T. (1977): *Associative memory: A system theoretical approach*. N.Y.: Springer.
- JONES, W. P. (1987): ANNA: An adaptive neural network associator for personal computer interfacing. En Proceedings of the IEEE First Annual International Conference on Systems, Man and Cybernetics. San Diego.
- LASHLEY, K. (1950): In search of engram. En *Psychological mechanisms in animal behavior*. Symposia of the Society for Experimental Biology. N.º 4. Academic Press.
- LURIA, A. R. (1966): *Higher cortical functions in the man*. N.Y.: Basic Books.
- MARR, D. y POGGIO, T. (1976): Cooperative computational of stereo disparity. *Science*, 194, 283-287.
- MASSARO, D. W. (1988): Some criticism of connectionism models of human performance. *Journal of Memory & Language*, 27, 213-234.
- MASSARO, D. W. (1989): Paradigms of psychological inquiry and process versus connectionism models of human performance. Paper presentado en 29 Meeting Anual de la Psychonomic Society. Atlanta.
- MCCLELLAND, J. L. (1976): Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 80-91.
- MCCLELLAND, J. L. (1988): Connectionism models and psychological evidence. *Journal of Memory & Language*, 27, 107-123.
- MCCLELLAND, J. L. y ELMAN, J. L. (1986a): The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.
- MCCLELLAND, J. L. y ELMAN, J. L. (1986b): Interactive processes in speech perception: The TRACE model. En McClelland y Rumelhart (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2*. Mass: MIT Press.
- MCCLELLAND, J. L. y RUMELHART, D. E. (1985): Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159-188.
- MCCLELLAND, J. L. y RUMELHART, D. E. (eds.) (1986): *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2*. Mass: MIT Press.
- MCCLELLAND, J. L. y RUMELHART, D. E. (1986a): A distributed model of human learning and memory. En McClelland y Rumelhart (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2*. Mass: MIT Press.
- MCCLELLAND, J. L. y RUMELHART, D. E. (1986b): Amnesia and distributed memory. En McClelland y Rumelhart (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2*. Mass: MIT Press.
- MCCLELLAND, J. L. y RUMELHART, D. E. (1988): *Explorations in parallel distributed processing*. Mass: MIT Press.
- MCCLELLAND, J. L., RUMELHART, D. E. y HINTON, G. E. (1986): The appeal of parallel Distributed Processing. En Rumelhart y McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1*. Mass: MIT Press.
- MCNAUGHTON, B. L. y MORRIS, R. G. M. (1987): Hippocampal synaptic enhancement and information storage within a distributed memory system. *TINS*, 10, 10, 408-414.
- MCCULLOCH, W. S. y PITTS, W. (1943): A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5, 115-133.
- MEYER, D. y SCHAVNEVELDT, R. (1971): Facilitation in recognizing pairs of words: Evidence of a dependence in retrieval operations. *Journal of Experimental Psychology*, 90, 277-234.
- MINSKY, M. L. y PAPERT, S. (1969): *Perceptrons*. Cambridge, Mass: MIT Press.
- MOZER, M. (1987): RAMBOT: A connectionism expert system that learns by example. En Proceedings of the IEEE First Annual International Conference on Systems, Man and Cybernetics. San Diego.

- NEELY, J. H. (1990): Semantic priming effects in visual word recognition: A selective review of current findings and theories. En Besner y Humphreys (eds.), *Basic Processes in Reading: Visual Word Recognition*. Hillsdale, N.J.: LEA.
- NEWELL, A. (1980): Physical symbol systems. *Cognitive Science*, 4, 135-185.
- NORMAN, D. A. (1986): Reflections on cognition and parallel distributed processing. En McClelland y Rumelhart (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 2. Mass: MIT Press.
- ODEN, G. C. (1988): Why the difference between connectionism and anything else is more than you might think but less than you might hope. Paper presentado en *28 Meeting Annual de la Psychonomic Society*. Chicago.
- PALMER, S. E. (1987): PDP: A new paradigm for Cognitive Theory. *Contemporary Psychology*, 32, 925-928.
- PITARQUE, A., ALGARABEL, S., DASÍ, C. y RUIZ, J. C. (1989): Evaluación de los métodos conexionistas de simulación. Ponencia presentada en el *I Symposium Nacional de Metodología de las Ciencias del Comportamiento*. Salamanca, noviembre 1989.
- PYLISHYN, Z. W. (1984): *Computation and cognition: Toward a foundation for cognitive science*. Mass: MIT Press.
- RATCLIFF, R. (1990): Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Bulletin*, 97, 285-308.
- ROSENBLATT, F. (1962): *Principles of neurodynamics*. N.Y.: Spartan Books.
- RUMELHART, D. E. (1977): Toward and interactive model of reading. In S. Dornic (ed.), *Attention and Performance VI*. Hillsdale, N.J.: LEA.
- RUMELHART, D. E. y MCCLELLAND, J. L. (1982): An interactive activation model of context effects in letter recognition. Part II. *Psychological Review*, 89, 60-94.
- RUMELHART, D. E. y MCCLELLAND, J. L. (eds.) (1986): *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1. Mass: MIT Press.
- RUMELHART, D. E. y NORMAN, D. A. (1982): Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, 6, 1-36.
- RUMELHART, D. E. y ZIPSER, D. (1986): Feature discovery by competitive learning. En Rumelhart y McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1. Mass: MIT Press.
- RUMELHART, D. E., HINTON, G. y MCCLELLAND, J. L. (1986): A framework for parallel distributed processing. En Rumelhart y McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1. Mass: MIT Press.
- RUMELHART, D. E., HINTON, G. y WILLIAMS, R. J. (1986): Learning internal representations by error propagation. En Rumelhart y McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1. Mass: MIT Press.
- RUMELHART, D. E., SMOLENSKY, P., MCCLELLAND, J. L. y HINTON, G. (1986): Schemata and sequential thought process in parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2. Mass: MIT Press.
- SCHNEIDER, W. (1987): Connectionism: Is it a paradigm shift for Psychology? *Behavior Research Methods, Instruments & Computers*, 19, 73-83.
- SCHNEIDER, W. y OLIVER, W. L. (1988). Using declarative knowledge to guide skill acquisition in a connectionism/control architecture. Paper presentado en *28 Meeting Annual de la Psychonomic Society*. Chicago.
- SCHNEIDER, W. y SHEDDEN, J. (1989): A connectionism/control simulation model of attention processing and cortex micro-structure. Paper presentado en *29 Meeting Annual de la Psychonomic Society*. Atlanta.
- SEIDENBERG, M. S. y MCCLELLAND, J. L. (1989): A distributed developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- SEJNOVSKI, T. J. y ROSENBERG, C. R. (1987): Parallel networks that learn to pronounce English text. *Complex Systems*, 1, 145-168.
- SELFRIDGE, O. G. (1959): Pandemonium: A paradigm for learning. En *Symposium on the Mechanisation of Thought Processes*. London: Stationary Office.
- SMOLENSKY, P. (1986): Information processing in dynamical systems: Foundations of Harmony Theory. En Rumelhart y McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1. Mass: MIT Press.
- SMOLENSKY, P. (1988): On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1-74.
- WALTZ, D. L. y POLLACK, J. B. (1985): A strongly interactive model of natural language interpretation. *Cognitive Science*, 9, 51-74.
- WIDROW, B. y HOFF, M. E. (1960): Adaptive switching circuits. *IRE WESCON Conv. Record*, part. 4, 96-104.
- ZEIDENBERG, W. (1987): Modeling the brain: A neural network approach. *Byte*, 12, n.º 14, 237-250.