

LA REPRESENTACION DEL CONOCIMIENTO EN LOS SISTEMAS CONEXIONISTAS

Soledad BALLESTEROS

Departamento de Psicología Básica

Universidad Nacional de Educación a Distancia

Se pone de manifiesto la dispersión existente en la Psicología Cognitiva actual cuando trata de explicar cómo se representa, almacena y recupera el conocimiento debido a la gran variedad de formalismos representacionales propuestos. Se compara como se explica la representación en los modelos simbólicos y en los modelos conexionistas. Aunque los orígenes de los modelos conexionistas se retrotraen al trabajo de McCulloch y Pitts (1943), el verdadero precursor fue Rosenblatt (1962) con los perceptrones que eran redes capaces de aprender modificando el peso de sus conexiones. Los nuevos modelos conexionistas son modelos no simbólicos en los que los pesos de las conexiones determinan el funcionamiento de la red. Se concluye que a pesar del rápido desarrollo de los modelos conexionistas todavía es necesaria mucha más investigación antes de poder conocer con seguridad cómo el ser humano representa la información sobre el mundo.

Palabras clave: Representación; Modelos simbólicos; Modelos conexionistas; Representación de rasgos; Representaciones distribuidas; Propagación hacia atrás.

The representation of knowledge in connexionist systems. The large number of representational formalisms that intent to explain how mind codifies, stores and recovers knowledge is noticed. Symbolic models of mind mantain that the representation of knowledge is obtained through a system of simbols related by means of a complex syntax. According to connectionst models the representation of knowledge inside the artificial neural nets is produced from the simultaneous interaction of a large number of units highly inter-connected. The first major contribution to those models is due to McCulloch and Pitts (1943) but it was Rosenblatt who showed that perceptrons could be trained.

The new connectionist models are nonsymbolic models of the mind in which the weights of its connections determine the functioning of the network. It is concluded that much more investigation is needed before we can ascertain how information about the world is represented in mind.

Key words: Mental representation; Symbolic models; Connectionist models; Distributed representations; Back-propagation.

El principal problema que tiene planteado hoy la Psicología Cognitiva consiste en explicar cómo se representa el conocimiento (Amsel, 1989; Holland, Holyoak, Nisbett y Thagard, 1986; Marslen-Wilson, 1989; Paivio, 1986; Palmer y Kimchi, 1986; Smythe, 1988, 1989, etc.). El número de teorías que tratan de explicar cómo se codifica, almacena y recupera el conocimiento se ha multiplicado durante los últimos años dando lugar a una amplia variedad de formalismos representacionales diferentes: redes semánticas (Quillian, 1968; Collins y Quillian, 1969; Collins y Loftus, 1975), marcos (Minsky, 1975), sistemas de producción (Anderson, 1983; Newell, 1973), imágenes mentales (Kosslyn, 1980; Paivio, 1986; Shepard y Cooper, 1982); representaciones estructurales (Cooper, Schacter, Ballesteros y Moore, 1982; Schacter, Cooper y Delaney, 1990), etc. La falta de acuerdo entre los teóricos sobre este tema, clave para la Psicología, es una muestra de la dispersión conceptual existente en el campo aunque el gran logro de la Psicología Cognitiva ha consistido precisamente en haber demostrado la utilidad de postular un nivel de representación mental, o serie de constructos que sirven para explicar la percepción, memoria, solución de problemas o la representación léxica.

Palmer (1978) en un trabajo seminal ha distinguido entre la existencia del mundo real, la existencia del mundo mental (que supone la representación cognitiva del mundo real) y la existencia del modelo mental (o representación del mundo mental). En su análisis del problema Palmer ha señalado que el modelo mental, al ser un modelo del mundo mental, debe ser también un modelo del mundo real. En este contexto, una teoría representacional debería constituir a la vez la descripción apropiada del mundo mental y del modelo mental. Palmer ha distinguido además dos tipos de representaciones: *la representación intrínseca* y *la representación*

extrínseca. Una representación es intrínseca cuando emplea relaciones que tienen las mismas constricciones inherentes a la relación que representa, por ejemplo, una imagen. En un sentido extrínseco, la relación entre la representación y lo representado es totalmente arbitraria, como en el caso de una proposición.

Las representaciones dependen de los procesos que son los que determinan la relación existente entre los elementos del objeto representado. La distinción actual entre representación y procesos recuerda la distinción clásica entre estructura y función (Mayor, 1985; Paivio, 1986). El concepto más frecuente de representación hace referencia a una *estructura* sobre la que se ejecutan unos determinados procesos. A su vez, los *procesos* se refieren a las actividades implicadas en la utilización de la información almacenada en una determinada estructura o representación. Sin embargo, no siempre resulta sencillo distinguir entre estructura representacional y procesos psicológicos (Paivio, 1986).

LA REPRESENTACION EN LOS MODELOS SIMBOLICOS

Los modelos simbólicos mantienen que la representación del conocimiento se realiza a través de un sistema de símbolos. En esto coinciden las Ciencias de la Computación y la Psicología Cognitiva. Tanto los humanos como los ordenadores son sistemas capaces de manipular símbolos, por lo que ambos sistemas puedan considerarse análogos. Estos modelos se inspiraron en el intento de desarrollar un sistema matemático puramente formal, libre de todo contenido que cristalizó en la lógica matemática de Russell y Whitehead. Sin embargo, Gödel demostró que no podía alcanzarse plenamente el objetivo total de la formalización pura, aunque sí se comprobó su universalidad. El segundo teorema de Gödel

señala que no puede formalizarse en el mismo sistema una prueba sobre la consistencia de un sistema razonablemente rico y complejo (Wang, 1988). La analogía entre la máquina universal, que podía programarse para calcular cualquier función formalizada explícitamente impulsó la creación de modelos simbólicos computacionales en Psicología (Pylyshyn, 1989). El fundamento de estos modelos está en que tanto la mente como la máquina de propósito universal operan sobre representaciones, que no son otra cosa que códigos simbólicos. Ambos sistemas representacionales se organizan en torno a tres niveles básicos para la arquitectura computacional: a) *El nivel semántico* que explica por qué estos sistemas saben cuáles son sus objetivos; b) *el nivel simbólico* que codifica mediante expresiones simbólicas el contenido semántico y dicta las reglas que permiten manipular dichos símbolos; y c) *el nivel físico* que actúa de soporte material del sistema.

Las representaciones en estos modelos se entienden como sistemas semejantes al lenguaje constituidas por un gran léxico de símbolos relacionados mediante una compleja sintaxis (Fodor y Pylyshyn, 1988).

LA REPRESENTACION EN LOS MODELOS CONEXIONISTAS

Frente a esta interpretación que ha sido la tradicional de la Psicología Cognitiva y las Ciencias de la Computación, ha surgido en los últimos años una nueva arquitectura no simbólica (Rumelhart, 1989), como una rama de la inteligencia artificial y del modelo cognitivo, llamada *conexionismo* o procesamiento paralelo distribuido, cuya idea central consiste en considerar que la cognición se produce a partir de la interacción simultánea de una serie de unidades semejantes a neuronas que se encuentran altamente interconectadas. Estos modelos intentan superar algunos problemas surgidos

por los modelos computacionales como la mala actuación en tareas relacionadas con la satisfacción simultánea de una serie de restricciones “ligeras” (McClelland, Rumelhart y Hinton, 1986).

Los creadores del *conexionismo* aunque ofrecen una explicación de los modelos mentales basada en la idea de cómputo mental (coincidiendo en esto con los defensores de los modelos simbólicos computacionales), no aceptan que el sistema tenga que manipular símbolos, a la vez que consideran que el modelo del computador está bastante alejado del conocimiento que se tiene actualmente sobre el funcionamiento del cerebro ya que parece poco probable que el cerebro implemente programas que funcionen serialmente como propone el modelo computacional de la mente (Feldman Ballard, 1982). Si se tiene en cuenta que una neurona necesita unos cuantos milisegundos para ejecutar una función y que una tarea de un nivel elevado de complejidad puede realizarse en unos cientos de milisegundos, el cerebro únicamente podría pasar por unos cien pasos discretos para realizar la tarea (lo que Feldman, 1985, ha llamado “el programa de los 100 pasos”). Para que el cerebro pueda pasar a través de los miles de pasos de información codificada en un programa de ordenador, es necesario que funcione en paralelo.

El *modelo cerebral de la mente* defendido desde esta nueva perspectiva afirma que las funciones cognitivas se producen en redes que operan en paralelo (Lloyd, 1989). Estos modelos no creen en la necesidad del programa almacenado ni de un lenguaje interior constituido por la manipulación de símbolos. Se trata, por tanto, de *modelos no simbólicos* de la mente en los que la actividad surge de la fuerza de las conexiones entre las unidades del sistema, no de los símbolos codificados dentro del sistema, siendo los pesos de estas conexiones los que determinan el funcionamiento de la red.

Aunque en algunos modelos conexionistas, como el modelo de activación interactiva para el reconocimiento de palabras propuesto por McClelland y Rumelhart (1981), cada unidad de procesamiento corresponde a un rasgo, una letra o palabra particular, por lo general, las representaciones en los sistemas conexionistas se encuentran *distribuidas* en el conjunto de unidades de la red. Esta red está dotada de las siguientes propiedades: 1) Tiene capacidad de autogenerarse, ya que nuevos "inputs" pueden clasificarse de la misma manera que otros semejantes existentes en el sistema; 2) tiene la capacidad de completar un patrón especificado de una manera incompleta, la red es capaz de proporcionar las activaciones correspondientes en los nodos que carecen de ellas; 3) tolera imperfecciones, de manera que cuando se lesionan ciertas unidades del sistema su funcionamiento se resiente ligeramente pero no se imposibilita totalmente. Esta última característica hace que estos modelos se asemejen más a la arquitectura cerebral, en la que el deterioro de una serie de neuronas, no impide el funcionamiento del sistema, sino sólo produce un deterioro que depende de la extensión y localización de la lesión.

Como en estas redes no existe un programa almacenado las representaciones se manipulan de manera no computacional. Para ello, se parte de un conjunto de representaciones que constituyen el "input" del sistema, y de otro conjunto que constituyen su "output"; no existiendo, sin embargo, representaciones intermedias. El desarrollo de esta forma diferente de entender la representación ha sido tan fuerte que para algunos se trata de un nuevo paradigma (Schneider, 1987).

LOS PRIMEROS MODELOS NEURONALES

Los orígenes de los modelos de redes neuronales se remontan al trabajo de Mc-

Culloch y Pitts (1943). Estos investigadores se sintieron cautivados por la semejanza existente entre los aspectos binarios de la lógica (en términos de verdadero o falso) de los interruptores eléctricos (con sus dos posiciones encendido o apagado) y las unidades neurales (con sus dos estados excitación o inhibición). Este parecido fue lo que les impulsó a aplicar los principios derivados de la lógica simbólica a la descripción de la actividad neuronal. McCulloch y Pitts mostraron que redes semejantes a neuronas eran capaces de realizar computaciones, de manera que el comportamiento de una célula nerviosa y sus conexiones con otras células podían representarse utilizando un modelo lógico en el que los nervios se equiparaban a los enunciados de la lógica y la propiedad de estar activados o desactivados eran semejante a las operaciones del cálculo proposicional con sus enunciados verdaderos o falsos. Su aportación supuso un gran avance en aquella época temprana porque hizo pensar en la posibilidad de formalización de las redes neuronales y abrió el camino que más tarde permitiría modificaciones del modelo. En aquellos primeros momentos, sin embargo, el principal problema consistía en determinar si las redes neuronales eran capaces de aprender (Rumelhart y Zipser, 1985; Cowan y Sharp, 1988a, 1988b).

Un avance importante en este sentido se produjo cuando Hebb (1949), haciendo uso de la idea expresada por Lashley (1942) sobre la *equipotencialidad* existente entre las regiones del cerebro, propuso que el aprendizaje se producía en las redes neuronales debido a un cambio en la fuerza de las conexiones entre los elementos de la red, mucho antes de que surgiera el conexionismo. Según Hebb, la conectividad del cerebro va cambiando continuamente con el aprendizaje, dando lugar a lo que se llamó "asamblea de neuronas". Esta propuesta de Hebb influyó en los creadores de redes adaptativas capaces de aprender aunque sus

ideas continuaron siendo meras especulaciones hasta que pudo simularse una de estas redes en un ordenador.

A pesar de la importancia de estas primeras aportaciones, el verdadero precursor de las redes neuronales capaces de aprender fue Rosenblatt (1962), que en su libro titulado *Principles of Neurodynamics*, expuso su trabajo sobre los *perceptrones*. Los perceptrones eran redes simplificadas capaces de aprender que estaban formadas por un conjunto de unidades sensoriales conectadas a otro conjunto de unidades motoras a través de un único estrato neuronal. Para entrenar un perceptrón se comenzaba asignando pesos arbitrarios a las conexiones aunque para obtener la respuesta deseada había que ajustar todas las conexiones. El procedimiento de Rosenblatt consistió en no producir ningún ajuste si la respuesta de la unidad era correcta; si la respuesta era incorrecta, se incrementaban los pesos de todas las sinapsis activadas si la unidad debía ser activada pero no lo estaba, y se disminuían dichos pesos cuando se obtenía el patrón opuesto. Tras múltiples ensayos, los pesos de las conexiones alcanzaban los valores deseados correspondientes a la computación del patrón meta. Lo importante de esta nueva concepción fue considerar al sistema de procesamiento como algo dinámico, interactivo, capaz de dirigirse a sí mismo sin necesidad de ser dirigido desde el exterior. Las redes de Rosenblatt eran capaces de aprender y no podían ser replicadas por los ordenadores basados en la arquitectura clásica. El problema estaba en que como Minsky y Papert (1968) demostraron los perceptrones sólo podían computar funciones lógicas sencillas. Los perceptrones, como sus descendientes las *adalinas* (Widrow y Hoff, 1960), o neuronas lineales adaptativas, cuya única diferencia con respecto a los perceptrones residía en el procedimiento de aprendizaje, estaban abocados a la temida explosión combinatoria, lo

mismo que sucedía en los modelos seriales (Papert, 1988).

Frente a estos problemas iniciales, por aquella época el computacionalismo clásico liderado por Newell y Simon (Newell, Shaw y Simon, 1958) había logrado que sus modelos simbólicos resolvieran teoremas lógicos utilizando el cálculo proposicional. Estos resultados favorables hicieron que prosperara la búsqueda de representaciones de tipo simbólico y que los logros alcanzados por los modelos paralelos cayeran en el olvido.

MODELOS CONEXIONISTAS ACTUALES: DIFERENCIAS CON LOS MODELOS SIMBOLICOS

Los modelos conexionistas actuales han resuelto satisfactoriamente el problema que las primeras redes neuronales no pudieron resolver. Este problema era su relativa incapacidad para encontrar la mejor solución pudiendo quedar atrapada la red en configuraciones metaestables antes de encontrar el verdadero mínimo global.

Hinton y Sejnowski (1983, 1986) han diseñado las Máquinas Boltzman que son redes adaptativas dotadas de unidades ocultas son capaces de encontrar la configuración estable de unidades activas e inactivas implementando el llamado procedimiento Monte Carlo. Estas máquinas son ya verdaderos instrumentos análogos en lugar de ser simples neuronas MacCulloch-Pitts con respuesta del tipo todo o nada.

Al igual que ocurre con los modelos simbólicos, el problema de la representación del conocimiento es el más importante que tiene planteado esta nueva perspectiva. En estos modelos cada entidad está representada por un patrón de actividad distribuida sobre una gran cantidad de elementos simples de computación, a la vez cada elemento de omputación está implicado en la representación de muchas entidades dife-

rentes. Como han señalado Hinton, McClelland y Rumelhart (1986): "El valor de este tipo de representación radica en la eficiencia con que hace uso de las habilidades de procesamiento de las redes de elementos simples de computación" (p. 77). La forma como los conceptos están representados dentro del sistema tiene importantes implicaciones para la actuación dinámica del mismo. En aquellos modelos en los que cada concepto está representado por una y sólo una unidad de procesamiento, un número determinado de unidades de procesamiento sólo pueden procesar el mismo número de conceptos, de una forma serial o paralela. Sin embargo, en el caso de algunos modelos conexionistas, todas las unidades de procesamiento que constituyen la red pueden participar en la representación de cada concepto, de manera que el sistema es capaz de representar muchos más conceptos, aunque sólo pueda activarse un concepto en cada ocasión (Neill y Klein, 1989).

Al igual que ocurre con los modelos clásicos, los modelos conexionistas son representacionistas. En este sentido, Rumelhart y McClelland (1986) insisten en que los modelos que ellos proponen se preocupan por estudiar la *representación* y los *procesos*. De igual manera, Smolensky (1988) considera que el conexionismo se enfrenta con regularidades a "nivel subsimbólico" e intenta proporcionar descripciones cognitivas construidas a partir de entidades que corresponden a los elementos constituyentes de los símbolos utilizados en el paradigma simbólico. Según Smolensky, el nivel subsimbólico tiene su propia semántica diferente de la del nivel simbólico o conceptual y así, las entidades que se representan en el paradigma simbólico mediante símbolos, en el paradigma subsimbólico se representan mediante un gran número de subsímbolos que participan en cálculos numéricos, no simbólicos (Smolensky, 1988, p. 3).

La diferencia principal entre estos modelos y los modelos cognitivos clásicos estriba en que aunque ambos asignan contenido semántico, el *PDP* lo asigna a los nodos mientras que los *cognitivistas* lo asignan a las expresiones simbólicas. Mientras los modelos *PDP* reconocen únicamente conexiones causales entre nodos, las teorías cognitivas lo hacen entre objetos evaluables semánticamente. En un modelo conexionista típico, el "input" del sistema viene dado por los valores de activación de las unidades "input" de la red. Estos valores numéricos suponen cierta *representación* del "input" que se propaga a lo largo de las conexiones hasta que aparece algún conjunto de valores de activación en las unidades del "output". Los valores de activación codifican el "output" que el sistema ha computado a partir del "input". Entre las unidades "input" y "output" puede haber otras unidades llamadas unidades ocultas que no participan en la representación ni del "input" ni del "output".

Un aspecto importante de los modelos conexionistas es que en ellos las representaciones están neurológicamente distribuidas correspondiendo sus estados de actividad a patrones de actividad neural. Esta concepción de la activación de las unidades y de la acción del sistema tiene un claro carácter neurofisiológico, de manera que la relevancia de la Neurociencia para los modelos *PDP* está bien documentada en la actualidad (Fernández Trespalacios, 1988; Nadel, Cooper, Culicover y Harnish, 1989). Los modelos conexionistas, frente al desinterés de los modelos cognitivos, defienden la necesidad de averiguar cómo es la estructura del cerebro para poder desarrollar modelos computacionales que se adapten a dicha estructura (Rumelhart, 1989), se trata por tanto de modelos más plausibles desde el punto de vista neurológico aunque antes de poder perfilar la relación existente entre estos modelos y la neurobiología hace falta

mucha más investigación (Nadel *et al.*, 1989).

Aunque se trata de sistemas más difíciles de manejar parecen más adecuados para apresar la fluidez de los procesos cognitivos humanos aunque no faltan quienes acusan a los modelos conexionistas de utilizar las ideas de la Psicología Cognitiva y la Inteligencia Artificial (Broadbent, 1985; Fodor y Pylyshyn, 1988). Otros, como Papert (1988), aseguran que las estructuras de los modelos conexionistas están basadas en problemas de juguete y carecen de un análisis teórico serio que asegure que el modelo vaya a funcionar cuando se reproduzca a tamaño real (lo que se conoce como problema del escalamiento).

En resumen, la representación en los modelos conexionistas es numérica en lugar de ser simbólica ya que las unidades de procesamiento de estos modelos se comunican unas con otras mediante señales numéricas en lugar de señales simbólicas. Las unidades de procesamiento suman los "inputs" de las conexiones con otras unidades de procesamiento de forma que cada "input" es modificado por la fuerza de la conexión (Churchland y Sejnowski, 1989). Este "output" es pequeño cuando los "inputs" están por debajo del umbral, pero va aumentando a medida que el "input" se incrementa como resultado del entrenamiento sucesivo de la red.

LA REPRESENTACION EN LOS MODELOS CONEXIONISTAS: ALGUNOS EJEMPLOS

A modo de ejemplo consideremos cómo se produce la representación en dos redes conexionistas. La primera, se trata del modelo más perfeccionado que existe hasta el momento, la llamada *NETtalk* (Sejnowski y Rosenberg, 1987), una red paralela que aprende a leer en alta voz. La segunda es una red sencilla desarrollada por Latimer y

Stevens (en prensa) que reconoce formas.

El sistema *NETtalk* está formado por tres capas compuestas por 309 unidades de procesamiento y 18.629 conexiones o pesos que deben ser especificados. Esta red no posee una organización preestablecida para que realice el procesamiento del "input" y logre el emparejamiento de las letras con sus sonidos correspondientes, sino que la estructura emerge espontáneamente durante el periodo de entrenamiento.

La primera capa de la red recibe como "input" las letras que constituyen una palabra, mientras que la última es la encargada de transformar el "input" en fonemas. La capa intermedia esta formada por unidades ocultas que al estar completamente conectadas con la primera y la tercera capa producen la transformación de las letras en sonidos.

Al principio la actuación de la red no es buena porque los pesos iniciales entre las conexiones se establecen al azar, pero en ensayos sucesivos estos pesos se van ajustando poco a poco en la dirección correcta con lo que mejora sustancialmente la actuación global de la red. La red aprende a realizar la transformación correcta entre texto y fonema utilizando el algoritmo de la "propagación hacia atrás", ajustando los pesos que controlan cada una de las conexiones hasta que finalmente llega a transformar las diferentes palabras en los fonemas correspondientes con bastante precisión.

La característica más destacada de esta red es que aprende de manera semejante a como lo hace el ser humano ya que: (a) tanto la red como el ser humano aprenden rápidamente pero el aprendizaje va decayendo con las repeticiones sucesivas; (b) ambos se benefician del aprendizaje distribuido o espaciado; (c) ambos son capaces de generalizar el aprendizaje a palabras nuevas con bastante precisión; y (d) La red muestra la llamada "degradación graciosa". Como ocurre con el cerebro humano, la

destrucción de una parte de la red no acaba con el funcionamiento del sistema, sino solamente produce una disminución del rendimiento.

¿Cuáles son los estilos de representación de esta red? Consideremos lo que ocurre en las diferentes capas: (1) La capa "input" funciona mediante representaciones locales ya que se utilizan unidades únicas para representar cada una de las letras del alfabeto; (2) cada fonema viene representado por un patrón de actividad producido entre las unidades correspondientes al "output", siendo la representación fonémica una representación distribuida. Sin embargo, cada unidad "output" viene codificada por un rasgo distintivo, o representación de rasgos. El concepto de representación distribuida se aplica en dos sentidos diferentes que se corresponden con la distinción comúnmente aceptada entre representación totalmente distribuida y representación de rasgos (Lloyd, 1989); y finalmente, (3) las unidades correspondientes a la capa oculta del *NETtalk* presentan una representación totalmente distribuida ya que los patrones de actividad están elegidos de una manera puramente arbitraria, de forma que la actividad desarrollada por una de estas unidades no tiene ningún significado aparente. Por tanto, no existe relación entre las unidades conceptuales que intervienen en el proceso cognitivo.

Si no existen reglas que indiquen la correspondencia entre el "input" y el "output" ¿cómo descubre la red la forma de pronunciar las palabras en una lengua tan irregular como el inglés?, ¿cómo puede generalizar y organizar el conocimiento? Midiendo el patrón de actividad desencadenado entre las unidades ocultas para "inputs" específicos, se ha podido comprobar que para cada patrón de letras que entran en el modelo existe un patrón de actividad entre las unidades ocultas. Normalmente para cada letra del alfabeto aparecen activadas en torno a 15 de las 80 unidades ocul-

tas, mientras el resto manifiestan poca o ninguna actividad. Este procedimiento se repitió con todas las letras de manera que se encontraron otros tantos vectores cada uno apuntando en una dirección en el espacio pluridimensional. Cuando se estudiaron las relaciones existentes entre vectores mediante el análisis "cluster" se encontró que los vectores correspondientes a las letras vocales se agrupaban juntos mientras que los correspondientes a las letras consonantes también se agruparon formando otro "cluster". Además, las letras que se pronuncian igual tendían a agruparse formando un mismo "clusters" (Chuchland and Sejnowski, 1989).

Este modelo, lo mismo que los perceptrones de capas múltiples y adalinas, es capaz de generalizar a una amplia variedad de tareas a la vez que es resistente al deterioro. Un *NETtalk* bastante dañado todavía es capaz de leer y hablar con una precisión de cerca del 40 por ciento y se recupera bastante bien cuando se le vuelve a entrenar. Estas propiedades propias de cualquier red dotada de representaciones distribuidas (Cowan y Sharp, 1988a). El modelo funciona muy bien aunque su capacidad es limitada cuando tiene que enfrentarse a la ambigüedad semántica y sintáctica.

Finalmente, vamos a considerar una sencilla red neural artificial que Latimer y Stevens (en prensa) han utilizado para simular el reconocimiento de formas geométricas y tratar de averiguar si existe evidencia de que los procesos de reconocimiento de patrones incorporados en la red neural poseen realidad psicológica. Esto es, si la forma de actuar de la red es semejante a la forma de actuar del perceptor humano. Para ello, entrenaron a una red consistente en una sola capa (perceptrón de capa única) constituida por 120 unidades de entrada, colocadas en una matriz de 10 x 12, y una sola unidad de salida, a activar la unidad de salida cuando se le presentaba la forma

“standard” y a desactivarla cuando recibiera uno de los patrones de comparación. El entrenamiento de la red se realizó mediante la regla de aprendizaje delta y la función de activación para la unidad output fue la suma ponderada de los inputs menos el umbral que se estableció en 0.5. La red fue capaz de realizar la tarea al cabo de 500 ciclos colocando más peso en las regiones más discriminativas de las formas. Los modelos conexionistas de percepción de la forma han recibido cierto apoyo en el trabajo de Latimer y Stevens (1992) al encontrar evidencia de que cuando los humanos aprenden a discriminar formas geométricas, asignan pesos de forma diferencial a las mismas partes componentes de las formas. Tanto estas redes como los humanos asignan mayor peso a las partes más discriminativas.

CONCLUSIONES

Las representaciones distribuidas gozan de importantes propiedades que no poseen las representaciones localizadas. En primer lugar, estas representaciones están dotadas de la propiedad de la *autogeneralización* ya que las representaciones con contenidos semejantes darán lugar automáticamente a patrones semejantes de activación. En segundo lugar, tienen la propiedad de poder *completar un patrón* ya que la red es capaz de rellenar las activaciones no existentes en algunas conexiones entre nodos. En tercer lugar, la red, a diferencia de los modelos computacionales simbólicos, es capaz de *tolerar pequeños errores* y de permitir que la actuación no sea totalmente perfecta.

Los sistemas conexionistas, como se ha podido apreciar a través de los ejemplos comentados, utilizan normalmente tanto de las *representaciones locales* como de las *representaciones distribuidas*. Mientras en la representación local cada unidad corresponde a una única representación, en la representación distribuida, cada representa-

ción esta constituida por la acción de un elevado número de elementos simples. La diferencia entre distintas representaciones no depende de las unidades que la formen sino del patrón de actividad manifestado por el conjunto de unidades, compartido por varias representaciones.

En definitiva, a pesar del rápido desarrollo de los modelos conexionistas en los últimos años queda todavía por determinar totalmente la suficiencia de estos modelos en el plano cognitivo. Como han señalado Nadel *et al.*, (1989), quizás el problema más importante al que tienen que enfrentarse estos modelos es explicar los procesos secuenciales que se dan en muchas de las actividades cognitivas humanas como son el lenguaje, la solución de problemas o el razonamiento, dentro del marco continuo y siempre dinámico propugnado por los conexionistas. Otro punto importante a considerar es que los sistemas biológicos tienen una estructura inicial que determina su forma de procesar los la estimulación ambiental. Las redes artificiales, al no tener esta estructura preestablecida, necesitan un periodo de entrenamiento inicial demasiado largo. Estos y otros muchos los temas necesitan todavía ser objeto de una profunda investigación antes de poder conocer con seguridad cómo el ser humano representa y procesa la información sobre el mundo que le rodea.

Agradecimientos

Parte de este trabajo fue realizado durante mi estancia como *Visiting Scholar* en la Universidad de Columbia (Estados Unidos) durante el curso 1989-90 gracias a una Ayuda del programa de Movilidad del Personal Investigador, Estancia en un Centro de Investigación Extranjero, concedida por la DGICYT. Agradezco a la Universidad de Columbia, y en especial a la Dra. L. A. Cooper, los medios que pusieron a mi alcance.

REFERENCIAS

- Amsel, A. (1989). *Behaviorism, neobehaviorism, and cognitivism in learning theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J. (1983). *The architecture of cognition*. Cambridge, MA: The MIT Press.
- Broadbent, D.E. (1985). A question of levels: Comments on McClelland and Rumelhart. *Journal of Experimental Psychology: General*, 114, 189-192.
- Churchland, P. S. and Sejnowski, T. (1989). Neural representation and neural computation. En L. Nadel, L. A. Cooper, P. Colicover, and R. M. Harnish (Eds.), *Neural connections, mental computation* (pp. 15-48). Cambridge, MA: The MIT Press.
- Collins, A.M. y Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Collins, A.M. y Quillian, R.M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- Cooper, L. A. , Schacter, D. L., Ballesteros, S. and Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory and Cognition*.
- Cowan, J.D. y Sharp, D.H. (1988a). Neural nets and artificial intelligence. *Daedalus*, 117, 85-121.
- Cowan, J.D. y Sharp, D.H. (1988b). Neural nets. *Quarterly Review of Biophysics*, 21, 365-427.
- Feldman, J. A. (1985). Connectionist models and their applications: Introduction. *Cognitive Science*, 9, 1-2.
- Feldman, J.A. y Ballard, D.H. (1982). Connectionist models and their properties. *Cognitive Science*, 6, 205-254.
- Fernández Trespalacios, J.L. (1988). El conexionismo. *Aldaba*, 11, 25-39.
- Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. En S. Pinker and J. Mahler (Eds.), *Connections and symbols* (pp. 3-71). Cambridge, MA: The MIT Press.
- Hebb, D.O. (1949). *Organization of behavior*. New York: Wiley.
- Hinton, G.E., McClelland, J.L. Y Rumelhart, D.E. (1986). Distributed representations. En D.E. Rumelhart y J.L. McClelland (Eds.), *Parallel distributed processing* (pp. 77-109). Cambridge, MA: MIT Press.
- Hinton, G.E. y Sejnowski, T.J. (1983). Optimal perceptual inference. Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 448-453.
- Hinton, G.E. y Sejnowski, T.J. (1986). Learning and relearning in Boltzmann machines. En D. E. Rumelhart y J.L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 282-317). Cambridge, MA: MIT Press.
- Holland, L.H., Holyoak, K.J., Nisbett, R.E. y Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge: MA: Mit Press.
- Kosslyn, S.M. (1980). *Image and mind*. Cambridge: Harvard University Press.
- Lashley, K. (1942). Persistent problems in the evolution of mind. *Quarterly Review of Biology*, 24, 28.
- Latimer, C. R. and Stevens, C. J. (en prensa). Eye movements measurements and connexionist models of form perception. En S. Ballesteros (Ed.), *Cognitive approaches to human perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lloyd, D. (1989). *Simple minds*. Cambridge, MA: MIT Press
- Marslen-Wilson, W. (1989). Access and integration: Projecting sound onto meaning. En W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp.3-24). Cambridge, MA: MIT Press.
- Mayor, J. (1985). Actividad humana y procesos cognitivos. En J. Mayor (Ed.), *Actividad humana y procesos cognitivos (Home-*

- naje a J.L. Pinillos), (pp. 3-36). Madrid: Alhambra Universidad.
- McClelland, J.L. y Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375-407.
- McClelland, J., Rumelhart, D. y Hinton, G. (1986). The appeal of parallel distributed processing. En D. Rumelhart y J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1* (pp. 3-44). Cambridge, MA: MIT Press.
- McCulloch, W.S. y Pitts, W. (1943). A logical calculus of the ideas immanent in neural nets. *Bulletin of Mathematical Biophysics*, 5, 115-137.
- Minsky, M. (1975). A framework for representing knowledge. En P.H. Winston (Ed.), *The psychology of computer vision* (pp.211-277). New York: McGraw-Hill.
- Minsky, M. y Papert, S. (1968). *Perceptrons*. Cambridge, MA: MIT Press.
- Minsky, M. y Papert, S. (1988). *Perceptrons*. Edición ampliada. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nadel, L., Cooper, L.A., Culicover, P. y Harnish, R.M. (Eds.) (1989). Introduction: Connections and computations. *En Neural connections, mental computation* (pp. 1-12). Cambridge, MA: MIT Press.
- Neill, W.T. y Klein, R.M. (1989). Reflexions on modularity and connectionism. En D.S. Gorfein (Ed.), *Resolving semantic ambiguity* (pp. 276-293). New York: Springer-Verlag.
- Newell, A. (1973). Production systems: Models of control structures. En W. G. Chase (Ed.), *Visual information processing* (pp. 463-526). New York: Academic Press.
- Newell, A., Shaw, y Simon, H.A. (1958). Effects of a theory of problem solving. *Psychological Review*, 65, 151-166.
- Paivio, A. (1986). *Mental representations*. New York: Oxford University Press.
- Palmer, S.E. (1978). Fundamental aspects of cognitive representation. En E. Rosch y B. Lloyd (Eds.), *Cognition and categorization* (pp. 259-303). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Palmer, S.E. y Kimchi, R. (1986). The information processing approach to cognition. En T.J. Knapp y L.C. Robertson (Eds.), *Approaches to cognition: Contrasts and controversies* (pp.37-77). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Papert, S. (1988). One AI or many? *Daedalus*, 117, 1-14.
- Pylyshyn, Z. W. (1989). Computing in cognitive science. En M.I. Posner (Ed.), *Foundations of cognitive science* (pp. 49-91). Cambridge, MA: MIT Press.
- Quillian, M.R. (1968). Semantic memory. En M. Minsky (Ed.), *Semantic information processing*. Cambridge, MA: MIT Press.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan.
- Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. En M. I. Posner (Ed.), *The architecture of cognition* (pp. 133-156). Cambridge, MA: The MIT Press.
- Rumelhart, D. E., McClelland J.L. and the PDP Research Group (1986). *Parallel distributed processing. Explorations in the microstructure of cognition: Vols. 1 y 2*. Cambridge MA: MIT Press.
- Rumelhart, D. E. y Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, 9, 75-112.
- Schacter, D. L., Cooper, L. A. y Delaney, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology: General*, 119, 5-24.
- Schneider, W. (1987). Connectionism: It is a paradigm shift for psychology? *Behavioral Methods, Instruments and Computers*, 19, 73-83.
- Sejnowski, T. y Rosenberg, C. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, 1, 145-168.
- Shepard, R.N. y Cooper, L.A. (1982). *Mental images and their transformations*. Cambridge: MIT Press.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1-74.
- Smythe, W.E. (1988). The semantic distinction

- between propositional and skillful knowing. *Journal of Intelligent Systems*, 1 (2), 135-184.
- Smythe, W.E. (1989). The case for cognitive conservatism: A critique of Dan Lloyd's approach to mental representation. *Behaviorism*, 17 (1), 63-73.
- Wang, H. (1988). *Reflections on Kurt Gödel*. Cambridge, MA: The MIT Press.
- Widrow, B. y Hoff, M.E. (1960). Adaptive switching circuits. *WESCON Convention Record*, 4, 96-104.