

EL ENFOQUE CONEXIONISTA EN PSICOLOGÍA COGNITIVA Y ALGUNAS APLICACIONES SENCILLAS EN DIDÁCTICA DE LAS CIENCIAS

Campanario, Juan Miguel

Aprendizaje de las Ciencias. Departamento de Física. Universidad de Alcalá

<http://www.uah.es/otrosweb/jmc>

juan.campanario@uah.es

Resumen. En este artículo se presenta el enfoque conexionista utilizado en psicología cognitiva para entender algunos procesos de comprensión y aprendizaje de las ciencias.

Palabras clave. Redes neuronales, conexionismo, aprendizaje, comprensión.

Summary. In this article I introduce the connectionist approach used in Cognitive Psychology in order to understand some comprehension and learning processes in science education.

Keywords. Neural networks, connectionism, learning, comprehension.

1. INTRODUCCIÓN

Uno de los problemas básicos en ciencia cognitiva consiste en entender los problemas de aprendizaje. Distintas teorías psicológicas sobre el aprendizaje nos han ayudado a explicar diversos dominios de fenómenos y datos relacionados con los problemas de aprendizaje y comprensión.

Quizá una de las teorías más exitosas en este dominio es la de los **esquemas** (McLelland y Rumelhart, 1986). Otros psicólogos han propuesto modelos de organización del conocimiento basados en estructuras organizadas tales como **esquemas** (Rumelhart y Ortony, 1977; Rumelhart, 1980); **guiones** (*scripts*) (Schank, 1980; Schank y Abelson, 1977); modelos mentales (Johnson-Laird, 1980) o **marcos** (Minsky, 1975). Según Graesser y Goodman, los esquemas son «paquetes naturales de conocimiento genérico altamente estructurado. Los

esquemas corresponden a diferentes dominios del conocimiento y pueden incluir conocimientos sobre personas, objetos, acciones, secuencias de acontecimientos, etc.» (Graesser y Goodman, 1985, p. 115). Por ejemplo, todos los profesores tienen un esquema más o menos similar sobre la acción de *examinar*. En esta acción intervienen un examinador y uno o varios examinados. También suele haber algún tipo de prueba (oral, escrita, práctica...).

En la última década, la teoría de los esquemas ha derivado en un nuevo enfoque a la ciencia cognitiva, el **conexionismo**. Este enfoque multidisciplinar, basado en el uso de **redes neuronales**, mantiene relaciones con campos tan aparentemente ajenos a la didáctica de las ciencias como son las matemáticas, la inteligencia artificial, la epistemología, la lingüística, la informática, la neurobiología y otros.

El enfoque conexionista se ha traducido en nuevos conocimientos y realizaciones prácticas. Por ejemplo, algunos programas informáticos de reconocimiento de voz están basados en el uso de redes neuronales que pueden adaptarse tras un proceso repetido de entrenamiento y reconocer la voz de diferentes personas. Además, existen sistemas basados en el uso de redes neuronales que interpretan síntomas de enfermedades o conceden créditos bancarios con más eficacia y menos errores que un experto. Otros sistemas basados en el uso de redes neuronales tienen aplicación para clasificar unidades o elementos de acuerdo con criterios de semejanza que son posteriormente interpretados por el investigador. En un trabajo anterior, por ejemplo, hemos utilizado una red neuronal para obtener mapas de la ciencia (Campanario, 1995).

A pesar del aumento de trabajos publicados en este nuevo campo de conocimientos, la aplicación de modelos conexionistas en didáctica de las ciencias es aún escasa. Sauleda y Martínez han comentado, de paso, este enfoque en un artículo sobre las propensiones esenciales en el escenario de la enseñanza de las ciencias experimentales (Sauleda y Martínez, 1994), pero no parece que la comunidad investigadora haya prestado una atención preferente al uso de redes neuronales para entender los problemas de aprendizaje de las ciencias. Ello no resulta raro si tenemos en cuenta que la teoría del esquema, precursora del enfoque conexionista, tampoco se emplea con asiduidad para entender los problemas cognitivos de aprendizaje de las ciencias.

Por otra parte, sabemos que los conocimientos previos desempeñan un papel fundamental en el aprendizaje y en la comprensión. Algunas teorías generales hacen de este hecho su punto clave, pero resultan de poca utilidad para explicar algunas *patologías* que los profesores consideran típicas de los alumnos de ciencias. Como veremos, los enfoques conexionistas pueden arrojar alguna luz sobre estas cuestiones.

De acuerdo con lo anterior, el objetivo general de este artículo es presentar a grandes rasgos los elementos esenciales del enfoque conexionista en psicología cognitiva para entender e interpretar algunos fenómenos y procesos cognitivos típicos del aprendizaje de las ciencias. Este objetivo está al servicio de una meta adicional: ampliar el dominio de conocimientos, técnicas y recursos explicativos del área de conocimientos de didáctica de las ciencias experimentales, en especial impulsar la aplicación de los modelos cognitivos para entender problemas de aprendizaje y enseñanza de las ciencias.

El resto del trabajo se organiza como sigue: en primer lugar se analizan los rasgos generales de las redes neuronales como instrumento básico de trabajo en el enfoque conexionista. Dado que existen muchos tipos de redes neuronales y de algoritmos de funcionamiento, resulta obligado centrar la discusión en algún modelo concreto. En el apartado 3 se describe un modelo sencillo basado en la competencia de las distintas unidades por la activación. También se discuten algunas limita-

ciones del conexionismo que conviene tener en cuenta. Los modelos sencillos, como el que presentamos, tienen aplicaciones interesantes que nos pueden permitir analizar y entender desde una sólida base teórica, y sin excesivas complicaciones matemáticas, algunos fenómenos y procesos del aprendizaje de las ciencias y esto se hace en la última parte del apartado 3. Dado que el dominio de conocimientos que abordamos es amplísimo, hemos elegido la intelegibilidad sobre la precisión. Por último se formulan unas conclusiones generales y se insiste en la necesidad de profundizar en la aplicación de nuevos enfoques en didáctica de las ciencias, pero siendo conscientes de las limitaciones de los mismos.

2. EL ENFOQUE CONEXIONISTA EN EL MARCO DE LA PSICOLOGÍA COGNITIVA

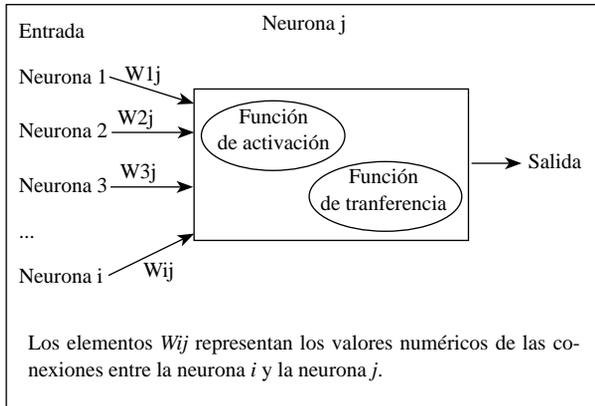
En la década de los noventa, muchos investigadores comenzaron a aplicar las redes neuronales a la psicología cognitiva. Este tipo de trabajos en redes neuronales configuran un punto de vista que se ha dado en llamar **enfoque conexionista** (McLelland y Rumelhart, 1989; McLelland y Rumelhart, 1986; Hanson y Burr, 1990; Hertz, Krogh y Palmer, 1991; Rumelhart, McClelland y el grupo PDP, 1992).

El enfoque conexionista en psicología cognitiva está basado en el cálculo mediante redes neuronales de determinadas propiedades que tienen su contrapartida psicológica. Una red neuronal está compuesta por un conjunto conectado de **neuronas artificiales**. Estas neuronas artificiales, en general, no existen físicamente y se implementan mediante programas de ordenador que almacenan en matrices de datos los parámetros que caracterizan la neurona, de la misma manera que almacenan y se tratan, por ejemplo, los datos correspondientes a las distintas posiciones físicas de un modelo climático que represente al planeta Tierra. Como se explica más adelante, el cálculo matricial permite gestionar un conjunto más o menos grande de estas neuronas artificiales que, como indicamos, son un mero artificio matemático.

Los modelos conexionistas se inspiran en la forma de trabajar en paralelo que tienen las neuronas del cerebro humano. Para ello se utilizan modelos basados en el paralelismo masivo, mediante los que se trata de modelar procesos cognitivos humanos con distinto nivel de complejidad (p.e., desde la percepción al pensamiento consciente). La interpretación sobre el sentido que debe asignarse a dichas unidades básicas originó ciertos debates en los primeros tiempos de la aplicación del enfoque conexionista a la psicología cognitiva (Smolensky, 1988; Fodor y Pylyshyn, 1988). Además, como se indica en un apartado posterior, algunos autores han señalado limitaciones importantes del conexionismo y de los modelos basados en redes neuronales.

A continuación se describen brevemente los elementos básicos que intervienen en una red neuronal (Fig. 1):

Figura 1
Elementos característicos de una red neuronal.



a) Un conjunto de n unidades de procesamiento (neuronas artificiales) que reciben impulsos de entrada de otras unidades y envían impulsos de salida a las restantes unidades o nodos. Cada nodo puede representar una variable, un rasgo, un concepto, etc. Como se ha indicado anteriormente, estas neuronas artificiales no tienen existencia física real y se implementan mediante programas de ordenador.

b) Un estado de activación de cada unidad (neurona), $a(n)$. Este estado de activación es función de las entradas que recibe la unidad y determina la salida que se envía a las restantes unidades. El estado de activación global viene dado por un vector de orden n y es la magnitud más importante de la red neuronal; de hecho, el vector final de activaciones es la magnitud que se suele comparar con los datos experimentales que se desea interpretar.

c) Un vector de orden n que representa los impulsos de salida de cada unidad. Estos impulsos de salida son función de la activación de cada una de las unidades.

d) Un patrón de conexiones entre unidades. Este patrón de conexiones puede representarse mediante una matriz de pesos o conexiones $W(n,n)$ que se multiplica por el vector de salida para obtener el vector de impulsos de entrada de cada unidad. Las conexiones pueden ser fijas o variables. En este último caso, las conexiones suelen depender del vector de activación.

e) Una regla de activaciones que combine, en un instante determinado, los impulsos de entrada en cada unidad junto con la activación actual de la unidad para obtener la activación en el instante siguiente.

De la descripción anterior se desprende una característica importante de los modelos conexionistas: todas las unidades están realizando cálculos, recibiendo entradas y enviando salidas simultáneamente (en paralelo). Las unidades son *tontas*, en el sentido de que realizan cálculos sin *saber* lo que están haciendo. Simplemente reciben impulsos, realizan unas operaciones matemáticas con

dichos impulsos y generan otros impulsos que van a parar a otras neuronas con las que están conectadas. Como consecuencia de ello y añadiendo, cuando resulte necesario, una regla de aprendizaje, es posible simular algunas de las características típicas del sistema cognitivo humano, tales como la degradación progresiva con el daño, el aprendizaje y generalización a partir de ejemplos, el procesamiento de información parcial, contradictoria o confusa y la actuación del cerebro en la resolución de ambigüedades (Rumelhart, McClelland y el grupo PDP, 1992).

Una de las aplicaciones más interesantes del conexionismo es simular el funcionamiento de los esquemas mediante redes neuronales (McClelland y Rumelhart, 1989), de ahí la enorme utilidad de este enfoque. Así, por ejemplo, diferentes elementos del vector de activaciones $a(n)$ pueden representar diferentes características o variables de un esquema y estas variables se pueden activar o no dependiendo de la interrelación o interconexión entre ellas. La activación final de cada elemento reflejaría la importancia o peso final de un nodo en una representación cognitiva. Si un conjunto de nodos está activado, y dado que cada nodo representa un concepto o una variable, se entiende que esos conceptos y variables son importantes en la representación final que tiene el sistema acerca de un dominio o una situación determinada.

Incluso problemas cognitivos que tienen que ver con la intencionalidad y que han preocupado durante años a los filósofos pueden implementarse en sistemas conexionistas (Falk, 1996). El paradigma conexionista ha dado lugar a una nueva disciplina basada en el uso de redes neuronales que promete, sin duda, influir en la orientación actual de la psicología cognitiva.

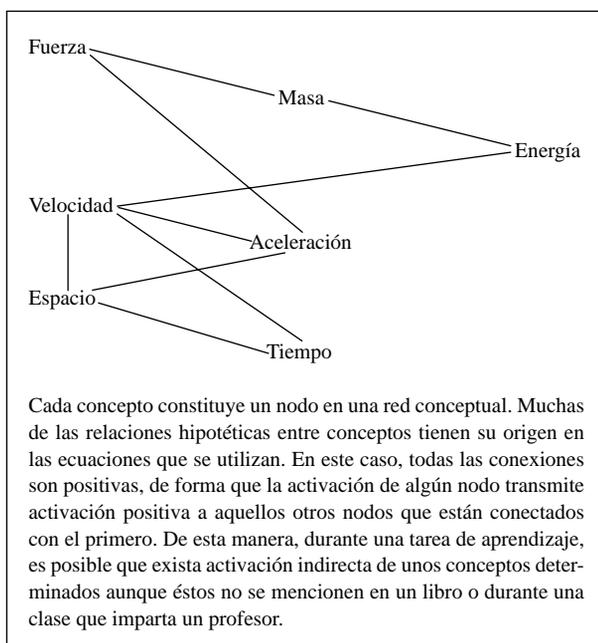
3. LA APLICACIÓN DE UN MODELO CONEXIONISTA PARA ENTENDER ALGUNOS PROBLEMAS Y PROCESOS DE ENSEÑANZA Y APRENDIZAJE DE LAS CIENCIAS

3.1. Un modelo conexionista sencillo basado en la competencia entre unidades

A continuación describiremos un modelo conexionista sencillo que aplicaremos posteriormente a algunos problemas de aprendizaje de las ciencias. En este modelo, los conocimientos del sujeto y , en su caso, la información que éste aprende se conciben como una red cuyos nodos se activan o desactivan mutuamente (Fig. 2). El conocimiento almacenado en la memoria a largo plazo del sujeto que aprende se concibe como una red que consta de conceptos. Las conexiones entre los elementos que componen la red pueden variar en magnitud dependiendo de las relaciones semánticas que existan entre los conceptos en cuestión. Además, estas conexiones pueden ser positivas o negativas. Las conexiones son positivas cuando los significados se refuerzan (p.e., fuerza y aceleración). Las conexiones son negativas cuando los significados se oponen (p.e., los distintos significados de una palabra tienden a desactivarse entre sí).

Figura 2

Representación hipotética de las relaciones entre conceptos de mecánica en un alumno que ha estado estudiando física.



Nuestra red consta, por tanto, de **n** nodos y supondremos ya formada la **matriz** de conexiones **W(n,n)** entre los nodos que representan el conocimiento del sujeto y, en su caso, la nueva información que se aprende o se procesa. El valor de las conexiones entre nodos deberá ser establecido por el investigador para reproducir los resultados experimentales que se desee simular, aunque deben evitarse los modelos con muchos grados de libertad porque resultan poco convincentes. En la práctica, es relativamente fácil encontrar un conjunto de parámetros que sirva para simular un dominio amplio de fenómenos, con lo cual la arbitrariedad aparente del sistema se reduce. Se suele utilizar un método de ensayo y error, tedioso pero eficaz.

La matriz anterior **W(n,n)** puede ser contradictoria o ambigua. Así, como se ha señalado, cuando existen conceptos que pueden tener distintos significados, existe una conexión negativa entre ellos. Dos conceptos contradictorios también están conectados negativamente. Estas conexiones negativas hacen que los nodos así conectados tiendan a desactivarse mutuamente. Por el contrario, cuando dos nodos están conectados positivamente tienden a reforzarse mutuamente.

Además, se define un vector columna **a(n)**, que especifica el valor de activación de cada uno de los **n** nodos que se procesan en el ciclo en cuestión. Dependiendo del proceso que se desee simular, se asigna a cada nodo un valor de activación (**1/n**), un valor cero a algunos nodos o un valor fijo a otros durante toda la simulación. Se

puede demostrar que, si los vectores de activación están normalizados (la suma o el módulo es 1), el proceso de multiplicaciones repetidas que se describe a continuación converge a un valor determinado.

Para simular el proceso de comprensión se deja que la activación se propague a toda la red multiplicando iterativamente la matriz de conexiones **W(n,n)** por el vector de activaciones para obtener un nuevo vector de activaciones de acuerdo con la expresión

$$a(j)^{t+1} = \sum_k w_{jk} * a(k)^t$$

donde los superíndices **t** y **t+1** expresan el valor de los componentes del vector de activaciones en dos iteraciones consecutivas. En determinados modelos en los que no tiene sentido un valor de activación negativo, estos valores se ponen a cero. Además, el vector de activaciones debe ser renormalizado después de cada iteración. Por ejemplo, se puede obligar a que el módulo del vector o la suma de sus componentes sea igual a la unidad, con el fin de asegurar la convergencia. El proceso iterativo de propagación de activaciones finaliza cuando dos vectores sucesivos difieren en una cantidad menor que un criterio, que se determina previamente.

Cada componente del vector de activaciones **a(n)** indica la importancia o el peso de cada nodo en la representación final. Por la propia dinámica de la multiplicación matricial repetida, los nodos con mayores interrelaciones entre sí habrán conseguido valores de activación mayores que los nodos menos conectados. En el cuadro I se ilustra con más detalle el funcionamiento del modelo que tratamos.

Cuadro I

Ejemplo de simulación de un proceso de activación y desactivación de nodos en una red neuronal.

En este cuadro se ejemplifica el funcionamiento del modelo general que se presenta en este trabajo. El modelo de Kintsch (1998) sigue, esencialmente, el mismo formato. Suponemos una matriz de conexiones entre nodos (conceptos) que representan conocimientos de un sujeto sobre un tema. La matriz de conexiones C quedaría del siguiente modo:

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N0
N1	a	b
N2	b	a	b
N3	.	b	a	b	.	.	b	.	.	c
N4	.	.	b	a	b	b
N5	.	.	.	b	a	.	.	c	.	.
N6	.	.	.	b	.	a	.	.	c	.
N7	.	.	b	.	.	.	a	.	.	c
N8	c	.	.	a	.	.
N9	c	.	.	a	.
N0	.	.	c	.	.	.	c	.	.	a

Los valores de conexión nulos han sido marcados con un punto. Se encuentran en la intersección de filas y columnas que corresponden a conceptos que no tienen relación entre sí. Los parámetros **a** y **b** corresponden, respectivamente, a conexiones de nodos con sí mismos o a conexiones positivas de un nodo con otros nodos. El parámetro **c** corresponde a una conexión negativa (por ejemplo, porque dos conceptos tienen significados contrarios). En el ejemplo que se está considerando se utilizarán los valores 2, 1 y -1 para los parámetros *a*, *b* y *c*. De esta manera la matriz quedaría como sigue.

N1	2	1	0	0	0	0	0	0	0	0
N2	1	2	1	0	0	0	0	0	0	0
N3	0	1	2	1	0	0	1	0	0	-1
N4	0	0	1	2	1	1	0	0	0	0
N5	0	0	0	1	2	0	0	-1	0	0
N6	0	0	0	1	0	2	0	0	-1	0
N7	0	0	1	0	0	0	2	0	0	-1
N8	0	0	0	0	-1	0	0	2	0	0
N9	0	0	0	0	0	-1	0	0	2	0
N0	0	0	-1	0	0	0	-1	0	0	2

A continuación se permite que la activación se propague a todos los conceptos que componen la red. Elegimos un vector columna de activaciones con todos los componentes inicialmente con el mismo valor igual a $1/n$, donde **n** es el número de nodos que intervienen. Por tanto, en el ejemplo que se discute, el vector de activaciones quedaría con todos sus componentes iguales a 0,1. El siguiente paso consiste en multiplicar repetidamente la matriz de conexiones por el vector de activaciones y se normaliza dicho vector en cada iteración. La normalización se hace de tal forma que la suma de los componentes del vector de activaciones sea igual a la unidad. Este procedimiento de multiplicación y normalización se repite hasta que dos vectores de activaciones sucesivos difieren entre sí en una cantidad criterio pequeña elegida previamente. En el ejemplo que se está considerando, después de realizar las iteraciones, el vector final de activaciones quedaría como sigue, para un criterio que hacemos igual a $\epsilon = 1 \cdot 10^{-7}$:

[.073, .150, .234, .217, .106, .106, .114, .000, .000, .000]

Como puede comprobarse, existen elementos con activación nula debido a que han sido desactivados por otros más numerosos con los que están conectados negativamente.

Una ventaja evidente de este tipo de enfoques es que permiten obtener **predicciones cuantitativas** de los valores de activación para correlacionarlos con datos obtenidos experimentalmente sobre recuerdo o comprensión con el fin de contrastar la validez de las suposiciones iniciales que nos llevan a elegir los parámetros de conexión. Esta posibilidad nos permite dar un paso más allá de las simples interpretaciones cualitativas basadas en modelos que tienen mucho de palabrería y poco de matemáticas (p.e., «comprensión superficial», «no se relacionan los conceptos», etc.).

Aunque el esquema general que se ha presentado puede parecer excesivamente sencillo y simplificado, está en la base del modelo de comprensión de construcción-inte-

gración de Kintsch, que ha sido ampliamente utilizado en diversas simulaciones de procesos cognitivos, lógicamente, con algunos elementos y complicaciones adicionales que se evitan aquí (Elosúa, 2000; Kintsch, 1988; Kintsch, Welsch, Schmalhofer y Zimny, 1990; Kintsch y Welch, 1990; Kintsch, 1998; Kinstch, 2000).

Es evidente que, cuando se simula el procesamiento de información mediante una multiplicación sucesiva de una matriz por un vector, no se pretende sugerir que el cerebro humano desarrolle ese proceso. Más bien se reproduce, mediante esta estrategia matemática, un proceso de competencia por activación; de la misma manera que el comportamiento del ala de un avión en el seno de un fluido se estudia resolviendo un sistema de ecuaciones diferenciales sin que ello signifique, obviamente, que el ala resuelve un conjunto de ecuaciones diferenciales.

3.2. Algunas limitaciones del enfoque conexionista

Aunque el enfoque conexionista y los tratamientos basados en el uso de redes neuronales han alcanzado un desarrollo notable en los últimos años, los modelos conexionistas adolecen de algunas limitaciones que es preciso conocer para valorar adecuadamente las posibilidades de esta metodología.

En primer lugar, existen fenómenos relacionados con el aprendizaje que no resultan fáciles de implementar mediante el uso de redes conexionistas; por ejemplo, la generación de nuevas construcciones cognitivas que aportan significados más ricos y extensos a los conceptos ya conocidos. Otra situación similar se produce cuando se intenta simular procesos de carácter implícito o procedimientos complejos (no olvidemos que los contenidos procedimentales son fundamentales en la enseñanza de las ciencias). Por otra parte, muchos de los logros de los modelos conexionistas actuales tienen que ver con el reconocimiento de patrones, pero estos procesos son mucho más limitados que los procesos cognitivos implicados en la comprensión y aprendizaje complejo de conceptos, principios y teorías. Además, un concepto clave de los modelos conexionistas, el de activación, no tiene una traducción inmediata y unívoca en términos relacionados con procesos cognitivos. Por otra parte, los mecanismos y aplicaciones matemáticas que utilizan los modelos conexionistas están tan alejados de los enfoques clásicos en psicología que Fodor y Pylyshin, dos críticos bien conocidos, señalan con cierta ironía que «todas las razones que hay para pensar que los modelos conexionistas podrían ser verdaderos, son razones para pensar que pudieran no ser psicología» (citado en García-Madruga, 1992, p. 25).

El lector interesado en profundizar en las limitaciones de los modelos conexionistas puede encontrar análisis más detallados en Pozo (2001) y Vega (1998). También puede ser interesante analizar la introducción de la edición española de la obra de Rumelhart, McClelland y el grupo PDP, citada en la bibliografía (Rumelhart, McClelland y el grupo PDP, 1992).

3.3. La simulación de algunos procesos y problemas de aprendizaje de las ciencias mediante un modelo conexionista

En esta sección se aplica el modelo descrito en la sección 3.1 y el cuadro I para entender algunos problemas de comprensión y aprendizaje de las ciencias. Es preciso insistir en que el modelo que hemos descrito es ciertamente sencillo y sólo representa uno de los muchos posibles dentro del marco general conexionista, y que otros problemas de aprendizaje podrían requerir la utilización de enfoques diferentes, más complicados.

a) El aprendizaje de las ciencias como establecimiento de nuevas conexiones entre esquemas o conocimientos previos

Una de las quejas más comunes de los profesores es que los alumnos tienden a aprender las ciencias como un conjunto de conocimientos inconexos o de dominios aislados con escasa relación entre sí. El enfoque conexionista nos permite entender el desarrollo de esquemas más completos como el paso de una matriz con bloques bien diferenciados y sin conexiones entre sí a una matriz con bloques interconectados.

La primera matriz es como sigue y representa el conocimiento de un alumno relativo a dos dominios de conocimiento que, en principio, aparecen sin relación entre sí: los nodos N1-N7 y N8-N0.

N1	1	1	0	0	0	0	0	0	0	0	0
N2	1	1	1	0	0	0	0	0	0	0	0
N3	0	1	1	1	0	0	1	0	0	0	0
N4	0	0	1	1	1	1	0	0	0	0	0
N5	0	0	0	1	1	0	0	0	0	0	0
N6	0	0	0	1	0	1	0	0	0	0	0
N7	0	0	1	0	0	0	1	0	0	0	0
N8	0	0	0	0	0	0	0	2	1	0	0
N9	0	0	0	0	0	0	0	1	2	1	0
N0	0	0	0	0	0	0	0	0	1	2	0

Como puede comprobarse con una sencilla simulación matemática, si se activa cualquier elemento de alguno de los dominios de conocimientos (esquemas), sólo se activan los elementos correspondientes a ese bloque. Nótese cómo los elementos N8, N9 y N0 no se activan, dado que no existe ninguna relación entre este dominio de conocimientos y el dominio de conocimientos representado por los nodos N1 a N7. La mecánica de la multiplicación repetida y la falta de conexiones entre bloques da como resultado los vectores finales que se presentan a continuación.

Vector inicial

[1,0,0,0,0,0,0, 0,0,0]
 [1,1,0,0,0,0,0, 0,0,0]
 [1,1,1,1,1,0,0, 0,0,0]

Vector final

[.07,.15,.23,.22,.11,.11,.11, .00,.00,.00]
 [.07,.15,.23,.22,.11,.11,.11, .00,.00,.00]
 [.07,.15,.23,.22,.11,.11,.11, .00,.00,.00]

La interpretación de los resultados anteriores es sencilla: por más que insistamos en activar un esquema determinado (bloque superior izquierdo, primeros siete componentes del vector de activaciones), si no existe relación (conexión) con otro esquema (bloque inferior derecho), es imposible que éste se active. Esta situación hipotética podría producirse, por ejemplo, en el caso de un alumno que está estudiando química y, debido a un aprendizaje superficial, ignora que los equilibrios ácido-base son un caso especial de los equilibrios en general. Para este alumno no tiene sentido el cálculo de las concentraciones en el equilibrio de disociación de un ácido o una base. Mientras no se establece una conexión concreta entre estos dos dominios de conocimiento, el alumno no aplica lo que sabe de los equilibrios químicos en general al caso particular de los equilibrios ácido-base. Un profesor con experiencia sabrá, seguramente, identificar otras situaciones en el aula que se pueden entender utilizando el modelo que estamos considerando.

Ahora, por efecto del aprendizaje, se establece una conexión entre ambos esquemas o dominios de conocimientos (conexión N6-N9).

N1	1	1	0	0	0	0	0	0	0	0	0
N2	1	1	1	0	0	0	0	0	0	0	0
N3	0	1	1	1	0	0	1	0	0	0	0
N4	0	0	1	1	1	1	0	0	0	0	0
N5	0	0	0	1	1	0	0	0	0	0	0
N6	0	0	0	1	0	1	0	0	1	0	0
N7	0	0	1	0	0	0	1	0	0	0	0
N8	0	0	0	0	0	0	0	2	1	0	0
N9	0	0	0	0	0	1	0	1	2	1	0
N0	0	0	0	0	0	0	0	0	1	2	0

En este segundo caso, la activación de un nodo cualquiera lleva la activación de **todos** los nodos, vía conexiones indirectas. Este sencillo modelo nos permite, asimismo, interpretar el fenómeno de la reconciliación integradora de Ausubel como nodos que representan conceptos y conexiones entre estos conceptos. Recordemos que la reconciliación integradora, en términos de la teoría ausubeliana, consiste en el establecimiento de relaciones entre conceptos o dominios en conflicto o aparentemente sin relación.

Vector inicial

[1,0,0,0,0,0,0, 0,0,0]
 [0,0,0,0,0,0,0, 0,0,1]

Vector final

[.01,.02,.04,.08,.03,.14,.02,.18,.30,.18]
 [.01,.02,.04,.08,.03,.14,.02,.18,.30,.18]

En términos de sucesos observables, el modelo anterior nos permite reproducir de una manera sencilla lo que ocurriría cuando un alumno oye o lee algo que activa en él determinada información. En ambos casos se activa la información relacionada con aquélla que se lee o se oye, pero en la segunda matriz la activación se propaga a otros nodos de forma indirecta mediante la conexión entre los nodos 6 y 9.

b) Interferencia de conocimientos inadecuados

La figura 3 ha sido obtenida de un guión de prácticas completado por un estudiante. Como puede comprobarse, este alumno relaciona inadecuadamente el principio de Arquímedes con el sencillo procedimiento de medir el volumen de un sólido mediante un recipiente aforado. En general, los profesores atribuyen este tipo de errores a deficiencias incomprensibles propias de los estudiantes, sin intentar profundizar más en los mecanismos cognitivos que permiten explicar tales deficiencias. El modelo sencillo que acabamos de aplicar en el apartado anterior nos permite entender el origen del problema. Es posible que se active un esquema relativo al principio de Arquímedes vía conexiones indirectas siguiendo la secuencia:

volumen – recipiente – líquido – empuje – Arquímedes

Aunque este mecanismo pueda parecer sorprendente a primera vista, existen indicios indirectos que lo apoyan. Por ejemplo, un estudio reciente demuestra que en un tesoro de unas 30.000 palabras inglesas, se puede llegar de una palabra cualquiera a otra palabra cualquiera, mediante una media de tres saltos a través de otras palabras con las que existen relaciones (p.e., relaciones conceptuales) (Motter, De Moura, Dai y Dasgupta, 2002).

Figura 3

Ejemplo obtenido de un cuaderno de prácticas de laboratorio de física general.

Conclusiones

¿Qué otros métodos podríamos utilizar para determinar volúmenes?
Por ejemplo:

a) Si conocemos la densidad del material.
b) Si disponemos de un recipiente aforado con precisión.

a) Por la fórmula

$$d = \frac{m}{v}$$

Si conocemos la densidad y además conocemos la masa, el volumen se puede calcular.

$$v = \frac{m}{d}$$

b) El principio de Arquímedes, el cual dice que, si introducimos un objeto en un líquido, el volumen del objeto es igual al volumen del líquido desalojado.

El alumno que ha contestado la pregunta ha activado los nodos correspondientes al principio de Arquímedes, probablemente mediante un mecanismo indirecto.

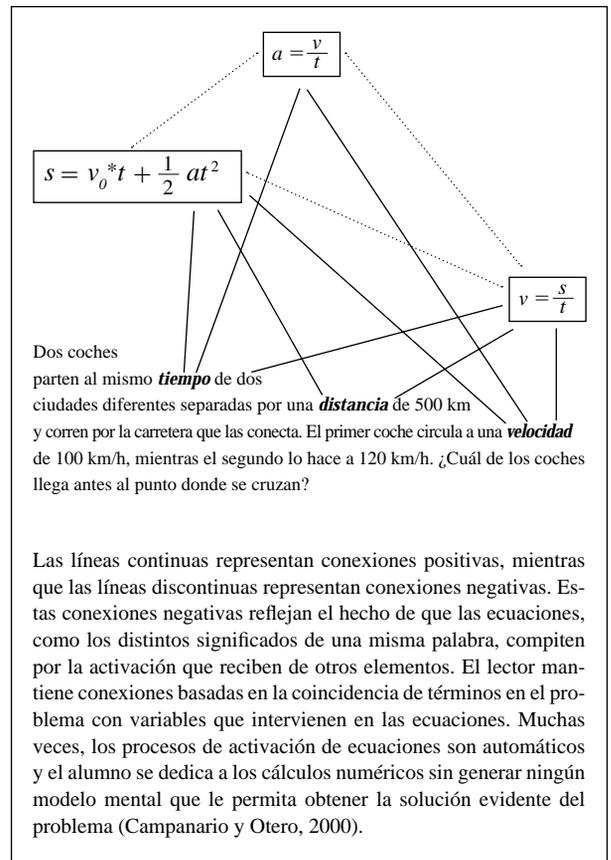
Si aplicamos un modelo de conexiones como el que se ha analizado anteriormente, es fácil entender este tipo de activación de conocimientos inadecuados. El mismo mecanismo explicativo nos permite entender la persistencia de las ideas alternativas y el hecho de que éstas con frecuencia se activen durante una tarea de aprendizaje o estu-

dio de un contenido científico, aunque no exista aparente relación entre dicho contenido y la idea alternativa. Así, por ejemplo, es posible que un alumno tenga asociada la idea de fuerza a la de dirección del movimiento, de tal manera que este último concepto se active automáticamente al activarse el concepto de *fuerza* (o viceversa). Otro caso similar puede producirse cuando los alumnos asocian sistemáticamente el volumen molar de 22,4 litros a cualquier sustancia, sólida, líquida o gaseosa (Furió, 1997). Una conexión inadecuada entre ambos esquemas puede estar en la base de la explicación de dicha patología.

Es importante insistir en que, según el modelo que aplicamos, la activación de conocimientos asociados a unos nodos determinados es un proceso **automático** que, en muchas ocasiones, está fuera del control del sujeto que aprende. Esto significa que las «patologías» típicas que muestran los alumnos no sólo son esperables, sino que resultan casi inevitables. Por ejemplo, en la figura 4 se presenta una interpretación de los procesos automáticos de activación de una ecuación determinada que es posible que desarrollen los alumnos cuando leen el enunciado de un problema de física a partir de los

Figura 4

Ejemplo de un posible mecanismo de activación automática de una ecuación almacenada en la memoria de un lector que procesa el enunciado de un problema de física.



datos y de la formulación superficial del problema. Este tipo de conductas ha sido identificado hace tiempo (Carrascosa y Gil, 1985) sin que hayamos sido capaces de profundizar en su comprensión utilizando modelos basados en la psicología del procesamiento de la información.

c) Medida de la consistencia del estado final de la representación alcanzada

Una aplicación adicional del enfoque que estamos tratando consiste en la medida de la consistencia del estado final que se alcanza tras un proceso de activación de conocimientos previos. Este problema está en la base de situaciones tales como la comprensión de un texto o el cambio conceptual, en las cuales el alumno debe decidir si una representación final es satisfactoria desde su punto de vista.

Podemos obtener una medida cuantitativa de la consistencia del estado final de la representación alcanzada mediante una magnitud propia de los modelos conexionistas que, a veces, se denomina **bondad de ajuste** o **armonía** (McLelland y Rumelhart, 1989). Por ejemplo, para el tipo de red que estamos utilizando, definimos la bondad de ajuste como

$$b = a^T W a$$

donde a y a^T representan el vector final de activaciones y su traspuesto. Recuérdese que C representa la matriz de conexiones. Nótese cómo de la definición anterior se desprende que la bondad es un valor numérico. ¿Qué sucede cuando se activan simultáneamente dos nodos (n_i y n_j) entre los cuales no existe conexión o existe una conexión negativa? La definición anterior da como resultado que la contribución a la suma final del producto $a_i * W_{ij} * a_j$ es nula o negativa y, por tanto, tiende a hacer disminuir la bondad de la red.

Aunque la definición anterior no es nueva, permite obtener una medida cuantitativa del grado de consistencia de una red conexionista. Así, por ejemplo, es evidente que una red con pocas conexiones negativas y muchas conexiones positivas entre sus elementos (lo cual representa un esquema o un dominio de conocimientos consistente) da lugar a un valor de bondad mayor que el que se obtendría para una red con conexiones negativas entre sus nodos o conceptos. En cambio, un dominio de conocimientos con escasas conexiones entre sí o con muchas conexiones negativas da lugar a un valor pequeño de bondad que correlaciona con la idea intuitiva de que es un dominio en el que uno *sabe poco y mal*.

d) Una interpretación del cambio conceptual

El cambio conceptual (entendido como el paso de un

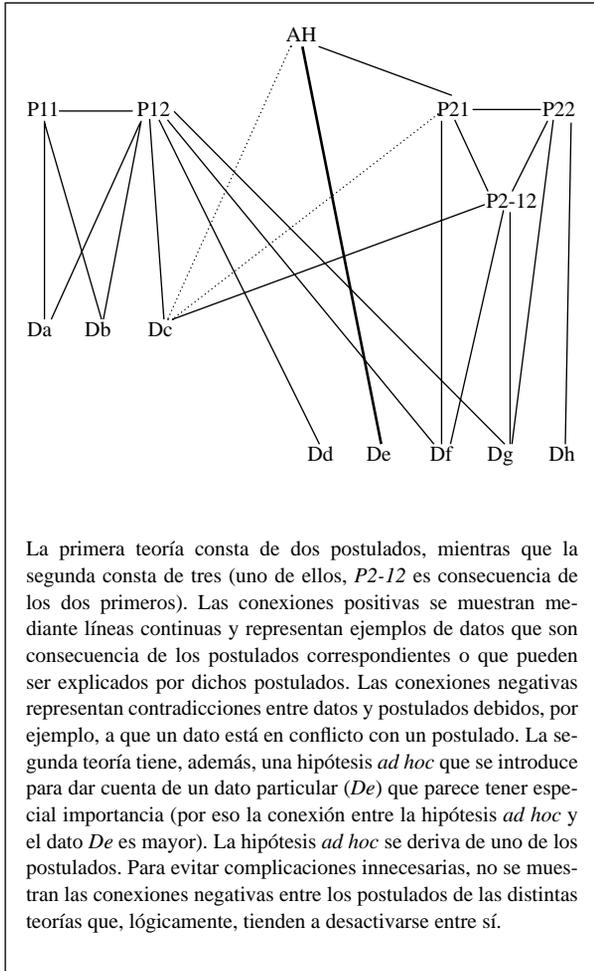
conjunto de ideas alternativas, casi siempre inadecuadas, a un conjunto de concepciones más acordes con las comúnmente admitidas por los científicos) ha sido uno de los objetivos tradicionales de la enseñanza de las ciencias (Marín, 1999). El cambio conceptual se inspira en el cambio desde una teoría científica a otra, un proceso bien estudiado en filosofía de la ciencia (Campanario, 2002).

Un modelo sencillo como el que estamos utilizando en nuestro trabajo puede ayudarnos a entender el papel de las explicaciones y las evidencias científicas en el proceso del cambio de teorías. Thagard ha analizado el cambio conceptual considerando las condiciones globales de coherencia de las teorías científicas. El elemento más llamativo del enfoque de Thagard es el uso que hace del formalismo matemático de redes neuronales para implementar la coherencia global de las teorías en competencia (Thagard, 1993). Esta coherencia depende tanto del número de hechos conocidos y nuevos que explican las teorías como de su grado de articulación.

La idea básica que subyace al enfoque de Thagard es la eliminación, en la representación final de la red (vector de activaciones), de una información por otra con la que la primera está conectada negativamente. Ésta es quizá la aplicación más evidente e inmediata del tipo de red que estamos estudiando y es una de las posibilidades más fructíferamente explotadas. De hecho, suele ser el ejemplo típico que se propone en los libros de texto introductorios (McLelland y Rumelhart, 1986; Hertz, Krogh y Palmer, 1991; <http://cogsci.uwaterloo.ca/Biographies/pault.html>).

En la figura 5 se presenta una ilustración del marco general que sirve para entender los procesos de cambio conceptual. En este esquema existen dos teorías compuestas por una serie de postulados (P_i) y datos experimentales que se derivan de dichos postulados o que se intentan explicar ($D_a, D_b...$). Existen conexiones positivas entre los postulados de cada teoría, mientras que las conexiones entre postulados de diferentes teorías son, lógicamente, negativas. Como puede comprobarse, el conjunto de datos que se explican con cada teoría refuerza en mayor o menor medida los postulados, dependiendo de la importancia que se conceda a cada dato. Si se genera una matriz conexionista y se permite que fluya activación entre los distintos elementos (datos y postulados) de la misma manera que hemos estado trabajando en las simulaciones anteriores, una de las dos teorías desplaza a la otra. Este desplazamiento depende tanto de la cantidad y variedad de datos que explica cada teoría como del grado de interconexión de los elementos y postulados que componen cada teoría. En el caso que analizamos, se puede evitar que una teoría muy compleja con poco apoyo empírico «derrote» a otra más sencilla con un respaldo superior concediendo una cantidad fija de conexión interna para ambas teorías de manera que esta cantidad fija se «reparta» entre los postulados que las componen.

Figura 5
Representación hipotética (y muy simplificada) de un dominio de conocimientos caracterizado por un conjunto de datos experimentales (Da, Db...) y dos teorías en competencia que intentan dar cuenta de algunos de los datos.



Un modelo similar, aunque algo más complicado, ha sido utilizado por Thagard (1993, 2000) para interpretar procesos históricos de cambios de competencia entre teorías científicas y competencia entre diversas argumentaciones científicas en términos de coherencia explicativa. No cabe duda de que estos modelos conexionistas pueden ayudarnos a entender que algunos científicos se resisten a abandonar un paradigma (entendido como un conjunto ricamente conectado de postulados, principios, observaciones y aplicaciones) incluso cuando existen evidencias en contra. A la vista del enfoque que analizamos, se hace evidente que conseguir un cambio conceptual no consiste únicamente en acumular hechos contrarios a un paradigma, por lo que cabe esperar resistencia ante los intentos destinados a modificar las teorías dominantes (Campanario, 2003; Campanario y Martín, 2003).

4. CONCLUSIONES Y PERSPECTIVAS

En general, en nuestra área de conocimientos prestamos poca atención a la fundamentación cognitiva de los procesos de enseñanza y aprendizaje de las ciencias mediante las teorías psicológicas del procesamiento de la información. Sirva, como ejemplo, un libro colectivo reciente que recoge las principales tendencias en la investigación en didáctica de las ciencias experimentales (Perales y Cañal, 2000). Como puede comprobarse por el índice de contenidos de este volumen, la aplicación de las teorías psicológicas del procesamiento de la información a nuestro dominio de conocimientos es limitada. Creemos, sin embargo, que las posibilidades de este enfoque en psicología y didáctica son enormes. Los sencillos ejemplos analizados en nuestro trabajo demuestran que podemos ir más allá de las interpretaciones tradicionales de algunos problemas bien conocidos en enseñanza y aprendizaje de las ciencias. En particular, nuestra opinión es que las fundamentaciones filosóficas e históricas deberían ir cediendo el paso, progresivamente, a aquellas otras basadas en los modelos psicológicos de comprensión y procesamiento de la información. No olvidemos que, en definitiva, el ser humano es un sistema complejo que procesa información procedente de muy diversas fuentes.

El modelo utilizado en este trabajo no es sino uno de los muchos que existen en el área de redes neuronales. No olvidemos que existen redes neuronales que «aprenden», entendiéndose el aprendizaje como un cambio en el conocimiento o, en el marco en el que nos movemos, como un cambio en las conexiones entre nodos. Existen diversos paradigmas de aprendizaje de redes neuronales (supervisado, no supervisado) y no cabe duda de que este formalismo puede ayudarnos en el futuro a interpretar y analizar más problemas de aprendizaje y enseñanza de las ciencias. Es posible, por ejemplo, utilizar este enfoque para simular los procesos de control de la propia comprensión en el estudio a partir de textos científicos (Otero y Kintsch, 1992), para simular el procesamiento de relaciones causales (Singer, 1996) o bien, en combinación con otro enfoque llamado *análisis semántico latente*, para medir la coherencia textual (Foltz, Kintschy y Landauer, 1998) o, incluso, la comprensión de metáforas (Kintsch, 2000).

La aplicación de los modelos citados puede ayudarnos a entender los procesos de aprendizaje de las ciencias de la misma manera que en física o en química recurrimos a explicaciones microscópicas o explicaciones basadas en modelos abstractos (p.e., transiciones de niveles protagonizadas por partículas puntuales cargadas que giran en torno a un núcleo) que nos ayuden a entender procesos o propiedades macroscópicas perceptibles (p.e., el color). Cuando recurrimos con éxito a estos modelos microscópicos, basados muchas veces en entidades ideales o inexistentes, nos sentimos con más derecho a concluir que hemos pasado de la **descripción** a la **explicación**. El paralelismo con los temas que hemos tratado es evidente: no podemos quedarnos satisfechos con la descripción de un problema de aprendizaje en términos de fallos aparentes de comprensión o de control de la comprensión.

Nos gustaría ir algo más allá y entender estos problemas en función de modelos que nos permitan formular algún tipo de predicción general basada en mecanismos básicos como son la activación o desactivación de la información.

En este trabajo hemos citado también algunas de las limitaciones del enfoque conexionista. Buscamos, con ello, evitar que este enfoque se considere una panacea o solución mágica que nos permita entender casi cualquier problema de aprendizaje de las ciencias. Además, en este trabajo centramos nuestra atención en las aplicaciones del conexionismo para modelar (en su ámbito de validez), fundamentalmente, los procesos de aprendizaje y dejamos sin abordar problemas relacionados con la enseñanza. Evidentemente, no podemos extrapolar los métodos que se siguen para enseñar a una red neuronal (un tema que aquí no hemos tratado) a la enseñanza de seres humanos. Aunque esta limitación es importante desde el punto de vista de la didáctica, no tiene excesiva importancia desde el punto de vista del desarrollo de la ciencia cognitiva.

El conocimiento y aplicación del enfoque conexionista exige un esfuerzo adicional de formación a los investigadores en didáctica de las ciencias. Este aumento en la exigencia de formación es el que cabe esperar de una disciplina que poco a poco se consolida a partir de unos inicios originados por los problemas de aprendizaje y la enseñanza de las ciencias y que extiende progresivamente su ámbito propio de conocimientos.

AGRADECIMIENTOS

Quiero expresar mi agradecimiento a José C. Otero, de la Universidad de Alcalá, que me ayudó allá por 1989 a descubrir la existencia del enfoque conexionista, y a Mike Mozer, de la Universidad de Colorado en Boulder, que me enseñó a utilizar las redes neuronales. Los comentarios de dos revisores anónimos ayudaron a mejorar este trabajo. Algunos de estos comentarios se han incluido en el artículo, especialmente en el apartado 3.3.

REFERENCIAS BIBLIOGRÁFICAS

- CAMPANARIO, J.M. (1995). Using neural networks to study networks of scientific journals. *Scientometrics*, 33, pp. 23-40.
- CAMPANARIO, J.M. y OTERO, J.C. (2000). La comprensión de textos de ciencias, en Perales, F.J. y Porlán, R. (eds.). *Didáctica de las Ciencias Experimentales*. Marfil: Alcoy.
- CAMPANARIO, J.M. (2002). *La enseñanza de las ciencias en preguntas y respuestas*. Alcalá: Universidad de Alcalá. Disponible en <http://www.uah.es/otrosweb/jmc>
- CAMPANARIO, J.M. (2003). Rejecting Nobel class papers and resisting Nobel class discoveries (en revisión).
- CAMPANARIO, J.M. y MARTIN, B. (2004). Challenging dominant Physics paradigms (en revisión).
- CARRASCOSA, J., GIL, D. (1985). La metodología de la superficialidad y el aprendizaje de las ciencias. *Enseñanza de las Ciencias*, 3, pp. 113-120.
- ELOSUA, M.R. (2000). *Procesos de la comprensión, memoria y aprendizaje de textos*. Madrid: Sanz y Torres, SL.
- FALK, A. (1996). A connectionist solution to problems posed by Plato and Aristotle. *Behavior and Philosophy*, 24, pp. 1-12.
- FODOR, J.A., PYLYSHYN, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, pp. 3-71.
- FOLTZ, P.W., KINTSCH, W. y LANDAUER, T.K. (1998). The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes*, 25, pp. 285-307.
- FURIÓ, C. (1997). Dificultades procedimentales en el aprendizaje de la química: la fijación y la reducción funcionales, en *Aspectos didácticos de física y química*, 7, Zaragoza: ICE-Universidad de Zaragoza.
- GARCÍA-MADRUGA, J.A. (1992). Introducción a la edición española, en Rumelhart, D.E., McClelland, J.L. y el grupo PDP (1992). *Introducción al procesamiento distribuido en paralelo*. Madrid: Alianza Editorial.
- GRAESSER, A.C. y GOODMAN, S.M. (1985). Implicit knowledge, question answering and the representation of expository text, en Britton, B.K. y Black, J.B. (eds.) *Understanding Expository Texts*. Hillsdale, Nueva Jersey: Lawrence Erlbaum.
- HANSON, S.J. y BURR, D.J. (1990). What connectionist models learn: Learning and representation in connectionist networks. *Behavioral and Brain Sciences*, 13, pp. 471-518.

- HERTZ, J., KROGH, A. y PALMER, R.G. (1991). *Introduction to the theory of neural computation*. Redwood, California: Addison-Wesley.
- JOHNSON-LAIRD, P.N. (1980). Mental models in cognitive science. *Cognitive Science*, 4, pp. 71-115.
- KINTSCH, W. (1988). The role of knowledge in discourse comprehension: A Construction-Integration model. *Psychological Review*, 95, pp. 163-182.
- KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- KINTSCH, W. (2000). Metaphor comprehension: A computational Theory. *Psychonomic Bulletin & Review*, 7, pp. 257-266.
- KINTSCH, W., WELSCH, D., SCHMALHOFER, F. y ZIMNY, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29, pp. 133-159.
- KINTSCH, W. y WELSCH, D.M. (1990). The Construction-Integration Model: A framework for studying memory for text. *Institute of Cognitive Science: Technical Report 90-15*. Boulder, Colorado: University of Colorado at Boulder.
- MARÍN, N. (1999). Delimitando el campo de aplicación del cambio conceptual. *Enseñanza de las Ciencias*, 17, pp. 80-92.
- McLELLAND, J.L. y RUMELHART, D.E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, Cambridge, Massachusetts: MIT Press/Bradford Books.
- McLELLAND, J.L. y RUMELHART, D.E. (1989). *Explorations in parallel distributed processing*. Cambridge, Massachusetts: MIT Press/Bradford Books.
- MINSKY, M. (1975). A framework for the representation of knowledge, en Winston (ed.). *The psychology of computer vision*. Nueva York: McGraw-Hill.
- MOTTER, A.E., DE MOURA, A.P.S., DAI, Y.C. y DASGUPTA, P. (2002). Topology of the conceptual network of language. *Physical Review E*, 65.
- OTERO, J.C. y KINTSCH, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, 3, pp. 229-235.
- PERALES, F.J. y CAÑAL, P. (2000). *Didáctica de las ciencias experimentales*. Alcoy: Marfil.
- POZO, J.I. (2001). *Humana mente: el mundo, la conciencia y la carne*. Madrid: Morata
- RUMELHART, D.E. (1980). Schemata: The building blocks of cognition, en Spiro, R.J., Bruce, B.C. y Brewer, W.F. (eds.). *Theoretical issues in reading comprehension*. Hillsdale, Nueva Jersey: Lawrence Erlbaum.
- RUMELHART, D.E. y McLELLAND, J.L. y el grupo PDP (1992). *Introducción al procesamiento distribuido en paralelo*. Madrid: Alianza.
- RUMELHART, D.E. y ORTONY, A. (1977). The representation of knowledge in memory, en Anderson, R.C., Spiro, R.J. y Montague, W.E. (eds.). *Schooling and the acquisition of knowledge*. Hillsdale, Nueva Jersey: Lawrence Erlbaum.
- SAULEDA, N. y MARTÍNEZ, M.A. (1994). Evolución y simbiosis de las propensiones esenciales en el escenario de la enseñanza de las ciencias experimentales. *Enseñanza de las Ciencias*, 12 (2), pp. 246-253.
- SCHANK, R.C. (1980). Language and memory. *Cognitive Science*, 4, pp. 243-284.
- SCHANK, R.C. y ABELSON, R.P. (1977). *Scripts, plans, goals and understanding*. Hillsdale, Nueva Jersey: Lawrence Erlbaum.
- SINGER, M. (1996). Comprehending consistent and inconsistent causal text sequences: A construction-integration analysis. *Discourse Processes*, 21, pp. 1-21.
- SMOLENSKY, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, pp. 1-74.
- THAGARD, P. (1993). *Cognitive revolutions*. Princeton, Nueva Jersey: Princeton University Press.
- THAGARD, P. (2000). Probabilistic networks and explanatory coherence. *Cognitive Science Quarterly*, 1, pp. 91-114.
- VEGA, M. (1998). La psicología cognitiva: ensayo sobre un paradigma en transformación. *Anuario de Psicología*, 29, pp. 21-44.

[Artículo recibido en noviembre de 2002 y aceptado en agosto de 2003]

