

De la máquina de Turing a la «máquina» de Boltzmann: dinámica interactiva y fenómenos globales en redes conexionistas

Manel Viader
Jaume Arnau
Universidad de Barcelona

A partir del análisis y crítica de algunos de los presupuestos básicos de la orientación dominante en psicología cognitiva se discuten las posibilidades del planteamiento conexionista al cual atribuimos, en el marco de los distintos niveles explicativos posibles, el carácter de alternativa verdaderamente psicológica. Las argumentaciones centrales se basan en las propiedades supuestamente atribuibles a las representaciones (particularmente, su carácter compuesto, sistemático y abstracto) y en la valoración de las posibilidades de aproximación a esas características desde el punto de vista conexionista. Se valora la posibilidad de complementar el enfoque microestructural propio del conexionismo con el análisis del comportamiento global del tipo de redes implicadas. Esta dinámica global, analizable desde una perspectiva topológica a partir del concepto de estabilidad estructural, es susceptible de mostrar modificaciones cualitativas que pueden asociarse con algunos fenómenos estudiados clásicamente en la psicología del pensamiento y en otros ámbitos.

Palabras clave: *Psicología Cognitiva, fundamentos teóricos, redes conexionistas, estabilidad estructural.*

The possibilities of the connectionist approach, which we consider a real psychological alternative within the different possible explanatory levels, are discussed on the grounds of analysis and criticism of some of the basic premises of the main cognitive psychology approach. The main arguments are based on properties supposedly attributable to representations (specially their complex, systematic and abstract character) and on the possibilities to approach these characteristics from the connectionist point of view. We consider the possibility to round off the microstructural approach, typical of connectionism, with the analysis of an holistic behaviour of all involved nets. Such overall dynamics, which can analysed from a to-

pologic view, starting from the concept of structural stability, can show qualitative changes that could be related to some phenomena studied classically in Thinking Psychology and other fields.

Key words: Cognitive Psychology, Theoretical Basis, Connectionist Networks, Structural Stability.

El panorama de la psicología cognitiva se ha visto alterado en los últimos años por la introducción de un conjunto de conceptos e incluso de aproximaciones globales que intentan plantear una alternativa a la orientación dominante, que podría ser caracterizada como «computacionalista-proposicionalista-sintáctica-discreta». Aunque las limitaciones de las aproximaciones habituales a los procesos cognitivos habían favorecido la búsqueda de aproximaciones paralelas de carácter esencialmente metodológico como es, por ejemplo, el caso de la neurociencia cognitiva, solamente en los últimos años puede hablarse de un cuestionamiento global de algunos de los supuestos de los modelos computacionales clásicos a partir, sobre todo, de la introducción en el marco de la psicología cognitiva de los denominados «modelos conexionistas» o de la llamada «perspectiva PDP» (Rumelhart y McClelland, 1986).

De acuerdo con Smolensky (1988a), lo que sostiene la aproximación conexionista es que algunos o muchos de los problemas de la psicología pueden abordarse mejor desde un nivel de análisis «subsimbólico» que desde la perspectiva «clásica» de la manipulación de símbolos; dicho de otro modo, que el nivel explicativo de la psicología, o al menos de muchos de los fenómenos a los cuales la psicología se enfrenta, no se halla en la manipulación de símbolos sino en el estudio del funcionamiento de redes de unidades elementales (no identificables con símbolos primitivos) que están sujetas a computación numérica.

La actitud con que se ha recibido este planteamiento entre los psicólogos cognitivos oscila entre dos posturas extremas que se concretan, bien en la afirmación de su irrelevancia o de su carácter no psicológico, bien en la consideración de su carácter auténticamente revolucionario, en el sentido kuhniano del término. La primera postura refleja el punto de vista de los teóricos que trabajan en los modelos computacionales que ellos mismos denominan «clásicos» (véase, por ejemplo, Fodor y Pylyshyn, 1988). Por su parte, los adeptos al nuevo enfoque conexionista defienden lógicamente su pertinencia y señalan sus importantes implicaciones en todos los niveles de la ciencia cognitiva.

En cualquier caso, la formulación conexionista nace en un contexto fuertemente interdisciplinar, tal y como sucedía con la orientación computacionalista «clásica», pero difiere de ésta tanto en los instrumentos conceptuales utilizados para la modelización de los procesos cognitivos como en el contexto científico general en el cual se enmarca. La apelación a la metáfora del ordenador es substituida por la atribución al sistema cognitivo humano de ciertas propiedades que hacen posible su tratamiento en el marco de la teoría de sistemas dinámicos. Resulta evidente que la introducción de este tipo de planteamientos suscita un conjunto de problemas metateóricos, teóricos y metodológicos de suma importancia, pero también lo es la amplia potencialidad teórica de los conceptos implicados

en esta orientación. Aunque el tratamiento de todos esos problemas no puede limitarse de modo alguno a un trabajo aislado, en las siguientes líneas intentaremos abordar algunas de estas cuestiones.

Algunas controversias básicas: niveles de análisis, poder representacional, composicionalidad, sistematicidad

En su artículo «Physical Symbol Systems», que se ha convertido ya en un verdadero clásico, Newell (1980) muestra el papel que puede jugar la teoría matemática de computación de símbolos en el marco de la ciencia cognitiva y articula las bases de lo que puede denominarse «paradigma simbólico» en psicología. Las ideas básicas del paradigma simbólico son bien conocidas: la realización de funciones cognitivas complejas exige que el sistema cognitivo sea capaz de representar y utilizar adecuadamente información estructurada; la forma más evidente e inmediata que puede tomar esta representación es la de un sistema de símbolos. Ese sistema debe incluir una serie de símbolos primitivos y un conjunto de reglas para su utilización y, en consecuencia, definir un determinado «lenguaje de la mente» cuya expresión abstracta más evidente es un formalismo de tipo proposicional. En conjunto, los procesos cognitivos pueden caracterizarse, desde esta perspectiva, como la operación de reglas sobre representaciones proposicionales. De este modo, las entidades básicas de funcionamiento del sistema cognitivo constituyen símbolos también (y sobre todo) en el sentido sintáctico del término, es decir, son susceptibles de manipulación a través de un mecanismo combinatorio que, a su vez, puede ser expresado en un conjunto de pasos computacionales discretos.

Obviamente, la tesis de que los procesos psicológicos se basan en un sistema simbólico es mucho más fuerte que la idea de que la entrada y/o la salida de un proceso sean en ocasiones patentemente simbólicas, como es el caso del uso del lenguaje natural. Se ha resaltado en muchas ocasiones que el hecho de que la conducta de un sistema pueda ser descrita en términos de símbolos y reglas de utilización de los mismos no significa que el sistema contenga verdaderamente un sistema simbólico (se trata de la vieja discusión sobre las representaciones del sujeto y las representaciones del psicólogo, o la que plantean los filósofos acerca de los fenómenos descritos por reglas y los fenómenos gobernados por reglas).

La conducta de cualquier sistema (cognitivo o no) puede ser descrita por referencia a un sistema simbólico reglado. Sin embargo, poco o nada se gana por el hecho de tratar, por ejemplo, el funcionamiento de una enzima como un operador de un sistema simbólico; no parece que a partir de esta descripción pueda llegarse a generalizaciones significativas que no puedan ser alcanzadas también por referencia a las propiedades físicas y químicas de dicha substancia. Parece que esa situación puede ser esencialmente distinta en el caso del comportamiento de sistemas de carácter cognitivo, puesto que la referencia a un nivel de análisis

físico-químico e incluso fisiológico no parece poder explicar, por ejemplo, la equivalencia de las respuestas cuando éstas son independientes de las propiedades físicas particulares del input. De ahí el necesario recurso al concepto de «representación», probablemente la idea central de toda la psicología cognitiva. En términos de Fodor (1975), existiría una reducción de señales (inscripciones físicas) a principios físicos, pero no se da una reducción semejante para los tipos de vehículos de representación (conjuntos de señales distintas que se refieren a un mismo símbolo).

Esta distinción es fundamental para cualquier enfoque no reduccionista de la psicología cognitiva. En efecto, es fácil darse cuenta de que la semántica de las representaciones no puede en sí misma causar la conducta (salvo que se recurra a la vieja «causalidad mental» del dualismo); solamente la forma o substrato material de la representación puede actuar como causa eficiente. Naturalmente, el problema que se plantea a continuación consiste en averiguar cómo se establece la relación entre las propiedades del estado físico subyacente a la representación y el contenido de la misma. En la respuesta a esta cuestión radica la idea central de la orientación «clásica»: lo que hace el cerebro es exactamente lo que hacen los ordenadores cuando computan funciones numéricas (Pylyshyn, 1986). La conducta de los ordenadores (y, supuestamente, también la de los seres humanos) es causada por las propiedades físicas de ciertas clases de sub-estados que corresponden a códigos simbólicos. Estos códigos deben reflejar todas las distinciones necesarias para hacer que la conducta corresponda a las regularidades expresadas en los términos semánticos.

En otras palabras, los símbolos son clases de equivalencia de propiedades físicas capaces al mismo tiempo de causar la conducta y de preservar ciertas interpretaciones y distinciones entre los términos semánticos. De acuerdo con la aproximación clásica, la consecución de este objetivo implica la existencia de una estructura sintáctica de las representaciones que pueda ser puesta en correspondencia con los rasgos semánticos, de modo que las propiedades semánticas de cualquier representación se reflejen en sus propiedades sintácticas. De este modo es posible construir sistemas físicos cuyos estados se modifiquen sistemáticamente en relación con la estructura sintáctica de sus representaciones o, en palabras de Fodor y Pylyshyn (1988), es posible construir una máquina gobernada sintácticamente cuyos estados (y transiciones entre ellos) satisfagan criterios semánticos de coherencia. En cualquier caso, resulta fundamental la idea de que la satisfacción de esos criterios y, en consecuencia, el carácter sistemático de la conducta del sistema, se produce en virtud del papel causal que se atribuye a la estructura sintáctica de las representaciones.

La crítica habitual a los modelos computacionales «clásicos» puede resumirse, con una muy considerable dosis de simplificación, en la afirmación de que se trata de modelos de la competencia (lo cual introduce inmediatamente dificultades en cuanto a su relación con la ejecución), sin criterios de auto-limitación y falsabilidad (puesto que se basan en un formalismo excesivamente potente que permite dar cuenta prácticamente de cualquier resultado experimental), con una significación psicológica discutible (dado el carácter estrictamente anti-psicológico del trabajo de los lógicos proposicionalistas) y con posibilidades de auto-compli-

cación y auto-organización muy limitados (salvo que se recurra a un aparato innato extraordinariamente complejo), de forma que parecen adecuarse muy poco al carácter básicamente adaptativo de la conducta. La existencia en nuestro contexto de trabajos excelentes que abordan esta problemática (por ejemplo, Rivière, 1984, 1986) nos exime, al menos momentáneamente, de extendernos excesivamente sobre estas cuestiones, aunque algo más adelante entraremos en algunos aspectos concretos que permitirán fundamentar el sentido de la alternativa conexionista y establecer los principales puntos de controversia entre ambas orientaciones.

Es necesario señalar que los modelos de tipo conexionista nacen con un problema que comparten con sus homólogos «clásicos» y, en general, con muchas de las formulaciones que se engloban en el marco de la psicología cognitiva. Nos referimos al origen extrapsicológico de algunos o muchos de los conceptos que se manejan. Es importante no perder de vista, en primer lugar, el hecho de que la perspectiva PDP surge en un momento en que se plantea con fuerza la necesidad de hallar un nuevo enfoque para el diseño de dispositivos computacionales, dadas las crecientes dificultades que encuentran los ordenadores digitales en la realización de gran número de tareas. Existe, por tanto, un «impulso tecnológico» subyacente al desarrollo de esta orientación, lo cual plantea inmediatamente el problema de la potencia explicativa («psicológicamente explicativa», por supuesto) de algunos de los conceptos implicados en ella. Ocurre, además, que se recurre a una «metáfora neuronal» de acuerdo con la cual el dispositivo de computación debería tener una estructura semejante a la del cerebro humano, es decir, estar organizado en forma de un conjunto de unidades elementales con un amplio nivel de conectividad. Resulta lógico, en consecuencia, el planteamiento de ciertas dudas sobre la significación psicológica de este tipo de modelos.

En relación con esta problemática se ha utilizado con frecuencia como instrumento conceptual la distinción entre niveles explicativos propuesta por Marr (1982). De acuerdo con esta distinción, y en términos de los más conspicuos representantes de la aproximación clásica, se trata de determinar si los modelos conexionistas pertenecen realmente al nivel propio de la psicología cognitiva o bien si se trata únicamente de modelos referidos a la implementación de dichos procesos en un dispositivo material concreto como es el cerebro humano (Fodor y Pylyshyn, 1988).

Es importante señalar, en primer lugar, que la utilización que se ha hecho en algunas ocasiones del concepto de niveles es ciertamente discutible. Por ejemplo, en su propio análisis, Broadbent (1985) intenta comparar los modelos localizacionistas clásicos con la perspectiva distribucional propia de muchos modelos conexionistas utilizando la conceptualización propuesta por Marr, pero lo hace a partir de un planteamiento como mínimo discutible. En concreto, Broadbent tiene en cuenta únicamente la existencia de dos niveles de explicación, a saber, el nivel de teoría de la computación y el de implementación material. Naturalmente, desde esta perspectiva la objeción básica respecto a los modelos distribucionales radicará en su equivalencia computacional con los modelos locales a los cuales pretende substituir. En particular, Broadbent recurre a un argumento clásico que, en última instancia, no tiene para nosotros significado alguno: los sistemas distribucionales no son capaces de llevar a cabo una computación que no

sea, a su vez, realizable por un sistema de almacenamiento «localizacionista» (Broadbent, 1985, p. 190). De modo más general, puede afirmarse que un sistema computacional clásico puede ejecutar cualquier operación que pueda llevar a cabo una arquitectura conexionista. A este respecto debemos señalar únicamente que dicha equivalencia computacional es débil y no guarda relación alguna con ningún criterio de verosimilitud psicológica. Como señala Rivière (1986), el recurso a la tesis de Turing puede ser suficiente para establecer la equivalencia a nivel computacional, pero es de poca utilidad para la psicología a causa de que ésta centra su interés en formas de equivalencia fuerte entre procesos psicológicos y los modelos que de ellos podamos construir.

El análisis de Broadbent se revela especialmente insuficiente si atendemos a la omisión por su parte de referencias específicas al nivel algorítmico propuesto por Marr. En realidad, el nivel algorítmico sería el más genuinamente psicológico, puesto que se plantea dar respuesta a las cuestiones relacionadas con la forma de representación de la información y con los procedimientos utilizados para su transformación. A nuestro entender, resulta bastante claro que los modelos conexionistas realizan propuestas alternativas y altamente específicas para los dos aspectos esenciales del nivel algorítmico: por una parte, la información es representada, de acuerdo con estos modelos, en forma de patrones de activación que admiten tanto propiedades de tipo todo o nada (generalmente a partir del establecimiento de umbrales) como de tipo continuo (las cuales, por cierto, son directamente compatibles con representaciones densas, tales como las imágenes mentales, y pueden aproximarse de forma muy eficiente al carácter analógico de este tipo de representación); por otro lado, la transformación de esas representaciones consiste en cambios en los patrones de activación en forma de modificaciones en la «fuerza» de las conexiones entre unidades de procesamiento y, por tanto, en el conjunto de interrelaciones que se establecen entre esas unidades. A nuestro entender, y tal como se verá más adelante, este tipo de arquitectura funcional permite, por ejemplo, recuperar la vieja y útil noción de fuerza asociativa sin caer necesariamente en las limitaciones de los modelos asociacionistas clásicos. En particular, será fundamental intentar establecer si este tipo de dinámica es capaz de construir representaciones compuestas y cuáles son las propiedades atribuibles a esas representaciones. Por otro lado, la información no es representada de modo localizado o de acuerdo con alguna «metáfora espacial», sino como resultado de la activación de múltiples unidades de funcionamiento, de modo que la información debe ser más «reconstruida» que «hallada». Parece evidente que el planteamiento de dos dinámicas de funcionamiento tan distintas debe poseer consecuencias empíricamente contrastables, aunque cuestión distinta son los problemas metodológicos que tal distinción implica.

Es interesante resaltar que, incluso a partir de modelos relativamente compactos (es decir, de carácter poco distribucional) es posible demostrar que el resultado empírico del funcionamiento de un sistema conexionista es muy distinto al que se obtendría a partir de una arquitectura funcional «clásica». Por ejemplo, en un sistema relativamente sencillo como es el NETL (Fahlman, 1979) no existiría relación entre el tiempo de búsqueda de un determinado contenido de memoria y el tamaño del conjunto o conjuntos a los cuales pertenece dicho ítem,

cosa que no ocurre, obviamente, en cualquier sistema de funcionamiento serial. Este ejemplo, al margen del valor concreto que pueda tener, es útil para replantear un tema de fondo como es el de la relación entre las funciones computadas por un sistema y la arquitectura funcional del mismo.

Tómese como ejemplo el de la máquina de Turing. Como es bien sabido, esta máquina es universal, en el sentido de que puede ejecutar cualquier función computable. Obviamente, la complejidad de la secuencia de operaciones ejecutadas por la máquina de Turing cambia de acuerdo con la tarea que debe realizar y, cosa mucho menos trivial, esta dependencia respecto a la tarea es distinta a la que puede establecerse en el caso de otro tipo de máquinas. Por ejemplo, el número de pasos exigidos para que una máquina de Turing lea un cierto número de símbolos presenta una relación cuadrática con la cantidad de símbolos implicada. En cambio, en una «máquina de registro», una arquitectura que posee lo que normalmente se denomina «acceso aleatorio» a la memoria del sistema, puede no existir relación entre número de símbolos y tiempo de lectura. En consecuencia, una arquitectura de este tipo puede ejecutar directamente ciertos algoritmos que, en cambio, no pueden ser llevados a cabo por la máquina de Turing salvo a través de la emulación de la máquina de registro, puesto que el repertorio de operaciones primitivas de cada una de ellas es distinto.

La distinción entre la realización de una computación por ejecución directa de cierto algoritmo mediante una determinada arquitectura funcional o por simulación de esa arquitectura en un sistema distinto posee implicaciones importantes. Por una parte, subyace a una de las más conocidas críticas al enfoque computacional: la confusión, difícilmente evitable, entre presuposiciones teóricas del modelo computacional y necesidades de programación del sistema (Johnson-Laird y Wason, 1977; de Vega, 1984). Parece claro que cuando se «desciende» desde el nivel de teorías de la computación al nivel algorítmico deben incorporarse necesariamente algunas presuposiciones relacionadas con la arquitectura funcional en la cual se desee implementar el modelo. Y, más importante aún, esas características de la arquitectura funcional *poseen consecuencias empíricas irreducibles*. Retomando el ejemplo anterior, una máquina de Turing que emule la arquitectura funcional de una máquina de registro seguirá mostrando, a diferencia de ésta, una relación cuadrática entre número de ítems a leer y tiempo de lectura (Pylyshyn, 1986). Naturalmente, y en coherencia con su propia postura, los teóricos «clásicos» presuponen que las restricciones impuestas por la arquitectura funcional del sistema (básicamente, en forma del establecimiento de operaciones primitivas) son equivalentes a la introducción de propiedades formales fijas en un sistema notacional. Globalmente, la tarea sería la de construir, una vez más, un sistema formal cuya gramática fuera equivalente a la de las operaciones cognitivas humanas y, en particular, que compartiera con éstas un cierto número de relaciones y propiedades primitivas. La idea fundamental es la de conseguir que el «nivel de agregación» de ambos sistemas sea semejante; es decir, se trataría de construir un sistema formal que ejecutara *directamente, y no por emulación de la arquitectura funcional*, las operaciones primitivas del intelecto humano. En esto radicaría el concepto de equivalencia funcional fuerte, puesto que la equivalencia de operaciones y relaciones primitivas garantizaría, al menos en gran parte, la se-

mejanza de las conductas externas (Pylyshyn, 1986, p. 107 y ss.). Aunque la discusión de este punto podría llevarnos muy lejos, permitásenos manifestar nuestro escepticismo en cuanto a la posibilidad de establecer, en el marco de los modelos «clásicos», los requisitos indispensables para la evaluación de la condición de equivalencia funcional fuerte (para una discusión más profunda de este problema, véase Rivière, 1986).

De acuerdo con todo lo anterior, nuestra hipótesis de partida respecto a la relevancia psicológica de los modelos conexionistas es que existe evidencia suficiente como para afirmar que pueden ser tratados como modelos psicológicos y que, como tales, pueden ser comparados con otros modelos del mismo nivel. Por citar solamente un ejemplo, McClelland y Rumelhart (1985) muestran que, en el campo del aprendizaje de conceptos, puede establecerse una comparación entre un modelo conexionista de tipo distribucional y la concepción prototípica de los procesos de categorización y conceptualización. En numerosas situaciones las predicciones que pueden realizarse a partir de ambos modelos son claramente divergentes. En particular, los modelos distribucionales son capaces de predecir las condiciones bajo las cuales la categorización es regulada por un prototipo o bien sigue una dinámica basada en ejemplares. Del mismo modo, un modelo distribucional puede ser comparado perfectamente con la teoría de los «logogenes» de Morton (1969, 1979, 1981) (con la interesante particularidad de que el modelo distribucional puede evitar la proliferación de «logogenes» para distintas variaciones de las condiciones de presentación), o con los llamados «modelos enumerativos» (Jacoby, 1983), sobre los cuales poseen ventaja en cuanto, por ejemplo, a la interpretación de la influencia del conocimiento anterior del sujeto sobre el almacenamiento de nueva información.

En todos estos casos, las ideas fundamentales de la orientación conexionista en cuanto a la representación y transformación de información parecen ser lo bastante poderosas y significativas como para derivar predicciones empíricas que, para muchas situaciones experimentales, divergen de las que pueden obtenerse a partir de otros modelos. A nuestro entender, ésta es una importante razón para defender el carácter psicológicamente significativo de los modelos de tipo distribucional en particular, y de los modelos conexionistas en general. Existen muchos otros ejemplos en la misma línea en los ámbitos de la percepción, la memoria, el lenguaje, la adquisición de conceptos, etc. (véase, por ejemplo, el libro fundamental —y casi fundacional, por su carácter sistematizador— de Rumelhart y McClelland, 1986), o la salida especial de *Cognitive Science*, núm. 9, 1985, etc.); por explicitar algunos casos recientes y muy distintos entre sí, podrían citarse las aplicaciones de los modelos conexionistas a temáticas tan dispares como el reconocimiento y pronunciación de palabras o, desde una perspectiva más global, el avance hacia la construcción de una «lingüística sub-simbólica» (Seidenberg y McClelland, 1989; van Orden, Pennington y Stone, 1990), fenómenos experimentales clásicos como el efecto Stroop (Cohen, Dunbar y McClelland, 1990) o la aproximación a ciertas patologías como la dislexia (Hinton y Shallice, 1991).

El tipo de argumentos que acabamos de citar no suele ser aceptado como suficiente por los teóricos «clásicos». Algunos autores (por ejemplo, Fodor y Pylyshyn, 1988; Fodor y McLaughlin, 1990) sostienen que ciertas características

del funcionamiento del sistema cognitivo y las propias capacidades de dicho sistema abogan de forma necesaria en favor de un sistema representacional de estructura composicional sobre el cual se definiría un conjunto de procesos mentales sensibles a tal estructura. De acuerdo con este punto de vista, si el conexionismo desea ser algo más que un simple asociacionismo (con todas las limitaciones que eso implica) debe aproximarse adecuadamente a esa estructura «compuesta» o compleja de las representaciones, y en ese caso debería recurrir necesariamente a arquitecturas funcionales «clásicas», de modo que su estatus sería únicamente el de implementaciones particulares (posiblemente verosímiles desde un punto de vista fisiológico) de ese tipo de arquitectura.

En realidad, lo que está en juego en estos argumentos es el problema de cómo explicar la sistematicidad del funcionamiento del sistema cognitivo (en el sentido trivial, por ejemplo, de que un organismo puede, en una situación de adquisición de conceptos, aprender a preferir un triángulo verde sobre un cuadrado amarillo, pero puede también aprender a preferir un círculo rojo por encima de un rectángulo azul, o puede construir con la misma facilidad la representación «Juan llama a Pedro» que la representación «Pedro llama a Juan»). Esto sugiere poderosamente la necesidad de definir elementos representacionales no dependientes de contexto, como los que subyacen a la sintaxis combinatoria propia de los modelos computacionales «clásicos». Ante esta situación, los teóricos conexionistas pueden probablemente recurrir a dos tipos de solución: o bien buscar una explicación no composicional de la sistematicidad a partir de representaciones dependientes de contexto, o bien aceptar el carácter composicional de las representaciones e intentar dar cuenta de los mecanismos de construcción de representaciones compuestas en el marco del funcionamiento de un sistema dinámico como el que ellos describen (pero no, por supuesto, a través de la implementación de una formalización simbólico-discreta de los principios computacionales en una red conexionista).

Smolensky (1989) realiza una propuesta en este último sentido, de acuerdo con la cual un sistema representacional de tipo distribucional podría explicar el carácter composicional de las representaciones a través del producto tensor de los vectores que representarían sus elementos constituyentes. Una crítica posible a este sistema y a otros del mismo tipo (memorias asociativas auto-recursivas, redes semánticas en paralelo, etc.) radica en la ineficacia causal de los componentes de las representaciones complejas, puesto que la activación de la configuración general no implica habitualmente la activación de dichos componentes. Este tipo de crítica ha producido cierta confusión entre algunos teóricos conexionistas. Por ejemplo, el propio Smolensky (1988b) señala la importancia de construir procesos sensibles a la estructura de las representaciones (lo cual implica un papel causal directo de dicha estructura sobre la conducta del sistema) y, al propio tiempo, afirma que las relaciones entre constituyentes no forman parte del mecanismo de procesamiento de la información en un modelo conexionista. Obviamente esta cuestión exige un mayor grado de reflexión.

Probablemente una respuesta adecuada al problema del papel causal de la estructura de las representaciones deba plantearse en la línea de profundizar en el significado del concepto de «representación compuesta» y en los mecanismos

que permiten la construcción de tales representaciones. En esta línea parece pertinente la distinción, propuesta por van Gelder (1990), entre «composicionalidad concatenativa» y «composicionalidad funcional», que parecen caracterizar, respectivamente, a las orientaciones clásica y conexionista. La composicionalidad concatenativa posee un carácter eminentemente sintáctico, tal y como se plantea en la orientación computacionalista clásica. En consecuencia, el significado de una representación compuesta se establecería a partir del significado asignado a sus constituyentes, vía reglas sintácticas de concatenación. En cambio, una representación compuesta no concatenativa (como, por ejemplo, la propuesta en el planteamiento de Smolensky, 1989) no está estructurada sintácticamente, puesto que no contiene señales de sus constituyentes primarios (aunque éstos en ocasiones pueden ser recuperables por descomposición). Esta distinción es importante a muchos niveles y es consecuencia prácticamente ineludible de las propiedades básicas de los modelos conexionistas. La esencia de la concatenación consiste, como se ha dicho, en la preservación de los constituyentes de la representación compleja, de modo que a cada uno de ellos le debe ser asignada una parte de los recursos de procesamiento del sistema. Por este motivo, para representar una estructura arbitrariamente compleja es necesario expandir los recursos también de forma arbitraria, algo que obviamente no puede hacerse en una red conexionista con un número finito de elementos ni tampoco, es importante remarcarlo, en los sistemas de procesamiento *reales*. De este modo, la crítica a la falta de memoria expandible en los sistemas conexionistas parece tener poco sentido, puesto que se ciñe a constatar el carácter limitado de tales sistemas. En cualquier caso, la alternativa conexionista a la expansión ilimitada de los recursos de procesamiento y almacenamiento radica en el carácter distribucional de las representaciones, que permite superponer diversas estructuras representacionales sobre las mismas unidades, al precio de destruir en el proceso los constituyentes elementales de tales representaciones.

Si se acepta el hecho de que pueden existir representaciones compuestas de carácter no sintáctico, ¿significa esto que la estructura de esas representaciones no puede jugar un papel causal en la conducta del sistema manteniendo al propio tiempo coherencia semántica? La respuesta a esta pregunta es totalmente afirmativa para los teóricos clásicos, que niegan la existencia de estructura interna en las representaciones propias de los modelos conexionistas, identificándolas pertinazmente con las representaciones inestructuradas propias del más ingenio de los asociacionismos (Fodor y Pylyshyn, 1988; Fodor y MacLaughlin, 1990). Sin embargo, Smolensky (1989) muestra que el hecho de rechazar el planteamiento ultra-local del asociacionismo clásico en favor de representaciones distribuidas permite aumentar de forma crucial la capacidad del sistema para la construcción de representaciones complejas, aunque no parece responder directamente a las objeciones planteadas por los teóricos clásicos en cuanto al papel de la estructura de las representaciones.

El problema que se plantea al conexionismo radica en la necesidad de definir en qué consiste y qué consecuencias posee el carácter «compuesto» de una representación generada, por ejemplo, a partir del producto de dos vectores referidos a representaciones más simples. La clave del funcionamiento del sistema

cognitivo radica en la posibilidad de reconocer representaciones con estructura similar o idéntica, cuya función en los procesos cognitivos sería también semejante, independientemente de la señal física concreta que sirviera inicialmente como soporte de la información. Abundando en la perspectiva conexionista, supongamos que los símbolos (incluyendo específicamente aquéllos que puedan considerarse como primitivos), a diferencia de lo que preconizan los teóricos computacionales «clásicos», posean algún tipo de «dinámica interna» y puedan definirse como un patrón de activación concreto de cierta red de unidades elementales. La cuestión se plantea entonces en los siguientes términos: ¿qué es lo que tienen en común los patrones de activación correspondientes, por ejemplo, a la imagen de un objeto y a la palabra que lo designa? (o a imágenes distintas del mismo objeto, o a palabras sinónimas, o a representaciones de nivel superior referidas a frases o a conceptos abstractos, etc.). O, al menos, ¿a qué nivel y con qué instrumentos conceptuales podemos abordar este problema desde una perspectiva distinta a la de los modelos clásicos?

Nuestra respuesta es la siguiente: Consideramos que las dinámicas subyacentes a ambas representaciones poseen una semejanza cualitativa que puede ser aprehendida mediante un análisis topológico, y que dicha semejanza se traduce a nivel funcional en la existencia de dinámicas comunes que interactuarían de forma idéntica con otros patrones de activación, al modo de resonancias características. Aunque este tipo de propuesta puede parecer relativamente abstracta (y en gran parte lo es, puesto que exige desarrollo ulterior) es necesario remarcar dos ideas iniciales: en primer lugar, la propuesta de un análisis de tipo topológico posee ventajas muy importantes, de entre las que citaremos únicamente, de momento, su capacidad para representar la complejidad de la información implicada en una representación (a través de la figura geométrica asociada a ella) y la posibilidad de establecer los estados estables del sistema y la potencial existencia de estados lo suficientemente inestables como para producir modificaciones profundas en su funcionamiento (en otras palabras, la posibilidad de estudiar su dinámica cualitativa); por otra parte, y precisamente por lo que acaba de indicarse, esta propuesta va más allá de las consideraciones que se realizan típicamente en el marco de la orientación conexionista y permite aproximarse a fenómenos cualitativamente importantes hasta ahora poco o mal explicados.

Desde nuestra perspectiva, la tarea que debe plantearse un estudio psicológico a nivel sub-simbólico es el análisis tanto de la cinemática como de la dinámica de este tipo de procesos, es decir, debe ocuparse tanto de la parametrización de los estados del proceso considerado como del estudio de su evolución temporal. Es importante resaltar, además, que las redes interactivas sobre las cuales se definirían estos procesos poseerían un carácter abstracto, de modo que un estado cualquiera de dicha red podría ser puesto en correspondencia con un estado fisiológico (mediante el correspondiente «mapa de correspondencias») o bien recibir una interpretación psicológica (por ejemplo, prueba de una hipótesis por parte de un sujeto que está intentando aprender un concepto). De este modo, el análisis del funcionamiento de este tipo de redes se revela como esencialmente «neutral» respecto al nivel explicativo adoptado (como lo es, en última instancia, cualquier instrumento matemático), y la función del investigador consiste en de-

finir en qué consisten y cómo pueden representarse mediante este tipo de instrumentos los estados (estables o no) del sistema bajo estudio, sean éstos de carácter psicológico o fisiológico. Lo que se nos ofrece, por tanto, es un instrumento de modelización cuyas potencialidades deben todavía evaluarse, y que resulta adecuado para el análisis de sistemas a los cuales se atribuyen determinadas propiedades dinámicas.

Una metáfora bastante ilustrativa del funcionamiento de las redes conexionistas se encuentra en el fenómeno de resonancia. De la misma forma en que dos osciladores pueden entrar, bajo ciertas circunstancias, en una dinámica común, dos conjuntos de unidades pueden entrar en resonancia siempre que presenten ciertas características particulares. Desarrollaremos algunas de estas cuestiones en el siguiente apartado de este artículo.

Resumiendo la línea de discusión que hemos seguido hasta aquí, creemos que el tipo de argumentos utilizado por los teóricos computacionales «clásicos» en su polémica con la orientación conexionista (la dificultad e incluso imposibilidad de aproximarse a las capacidades reales del sistema cognitivo desde modelos de este tipo) no resulta convincente al menos por dos razones: por un lado, por el hecho de que los propios modelos computacionales clásicos se muestran incapaces de aproximarse a dichas capacidades en condiciones razonables; por otro, como hemos señalado unas líneas más arriba, no está claro que las dificultades de los modelos conexionistas para dar cuenta de tales características y capacidades sean insalvables o se deban a limitaciones intrínsecas de dichos modelos. Al margen de las consideraciones ya expuestas tómese, por ejemplo, el caso del carácter productivo de muchos procesos cognitivos y, en particular, del lenguaje y del razonamiento. Al parecer, para los teóricos computacionales «clásicos» el carácter productivo de ciertos procesos cognitivos radica en la posibilidad de generar nuevas representaciones a partir de la aplicación de determinados operadores. Un ejemplo muy claro de este tipo de funcionamiento se encuentra en los sistemas de producción (por ejemplo, Anderson, 1983). El problema, obviamente, radica en cómo y cuándo se establecen las posibilidades algorítmicas del sistema cognitivo para una determinada tarea o conjunto de tareas. Se trata exactamente de la misma discusión planteada en torno al carácter generativo del conocimiento subyacente a la competencia lingüística y, a nuestro entender, no puede llevar sino a la misma conclusión final: la referencia última a un cuerpo de conocimiento declarativo y procedimental prefijado y de carácter innato. Dicho de otro modo, cuando la información utilizada por un sistema determinado está construida de modo que se ignora toda dinámica subyacente a los símbolos utilizados el «lenguaje de la mente» resultante no puede incrementar su poder expresivo; por el contrario, se ve obligado a poseer un conjunto pre-existente de predicados que determinan de forma definitiva el poder representacional de dicho sistema. La gran tarea de los modelos conexionistas consiste precisamente en estudiar la forma en que los propios símbolos, sus relaciones y transformaciones pueden ser entendidos como fenómenos emergentes de una dinámica no simbólica.

En cualquier caso, y sin entrar en la discusión de la hipótesis innatista propia de los modelos computacionales «clásicos», sí puede decirse, a nuestro en-

tender, que el hecho de postular un cuerpo de *conocimiento* innato cuyos límites puedan fijarse arbitrariamente, junto con la tradicional práctica de acomodar *ex post facto* la estructura de los modelos computacionales a partir de la observación del comportamiento real del sujeto, configuran una forma de trabajo bastante alejada de lo que tradicionalmente hemos llamado ciencia. O, expresándolo de otro modo, la retroacción infinita a la que conducen los problemas interpretativos de un lenguaje de la mente tal y como lo plantean los teóricos «clásicos» (y que tan poderosamente recuerda a la «mística» wittgensteiniana) no parece suministrar un marco excesivamente adecuado para la actividad teórica y empírica de los psicólogos cognitivos.

Dinámicas coherentes y emergencia de fenómenos en redes de tipo conexionista

La adopción de un enfoque microestructural, tal y como presuponen los modelos conexionistas, no está en contradicción con la existencia de fenómenos que se producen a un nivel mucho más holístico. Por el contrario, en este apartado intentaremos demostrar que la existencia de tales fenómenos es el resultado natural del funcionamiento de redes interactivas. De hecho, esta idea ha estado de alguna forma presente a lo largo del desarrollo de la psicología, ya sea en forma de leyes guesálticas, principios emergentistas más o menos difusos, etc.

Es preciso señalar que la simulación en ordenador de redes paralelas interactivas ha mostrado la existencia de efectos colectivos como reverberación difusa, fenómenos oscilatorios, transiciones de fases y, en general, un conjunto de fenómenos semejantes a los tratados en la emergente (y altamente interdisciplinar) teoría de las dinámicas caóticas. Este tipo de hallazgos sugiere la posibilidad de entroncar el estudio del tipo de modelos que proponen los teóricos conexionistas en el marco más general del análisis del comportamiento coherente de sistemas compuestos por múltiples unidades, tal como ocurre, por ejemplo, en biología, en química o en física.

El problema fundamental para el funcionamiento correcto de redes interactivas radica en evitar su «muerte entrópica», es decir, su aproximación a un estado «final» de equilibrio. Ahora bien, el alejamiento de la situación de equilibrio es al propio tiempo condición necesaria y consecuencia del funcionamiento coordinado de un conjunto de unidades que en situaciones próximas al equilibrio no mantendrían ninguna relación entre sí. Aun en situaciones próximas al equilibrio, cualquier sistema presenta ciertas fluctuaciones (por ejemplo, en el caso de un sistema químico, las fluctuaciones locales en las concentraciones, debidas al movimiento molecular; en un ecosistema, las fluctuaciones producidas por los cambios en los recursos energéticos disponibles, etc.); superada una distancia crítica respecto al estado de equilibrio, estas fluctuaciones pueden ampliarse y conducir al sistema a nuevos estados estables. Estos regímenes estables pueden mantenerse únicamente a través de un continuo intercambio de materia, energía y, añadimos nosotros, *información*, con el medio. La posibilidad de un análisis

a nivel de transmisión y transformación de información nos permite rechazar, al menos en principio, cualquier veleidad reduccionista. No se trata de encontrar un fundamento físico, químico, biológico o fisiológico que dé cuenta de cualquier morfología en la que se produzcan transiciones entre estados estables, sino de construir una teoría general de la estabilidad estructural aplicable a todo sistema que cumpla con un conjunto de requisitos mínimos. Prigogine (1983) afirma que esta tarea es posible y sugiere cuáles deben ser las características de un sistema al cual sean aplicables los conceptos que estamos analizando:

a) Debe tratarse de sistemas caracterizados por un comportamiento coherente de cierto número de unidades.

b) El problema de los límites juega un papel fundamental. Se trata siempre de sistemas no aislados, que intercambian materia, energía e *información* con el medio externo.

c) La evolución temporal de estos sistemas muestra, incluso en intervalos muy cortos, puntos singulares que marcan la aparición de fenómenos macroscópicos nuevos.

El tipo de análisis que estamos proponiendo ha sido aplicado a ámbitos tan distintos como el urbanismo, la construcción de modelos «dinámicos» en economía en oposición a los modelos lagrangianos, el comportamiento coherente de poblaciones de insectos, etc. A nuestro entender, los modelos conexionistas son claramente susceptibles de análisis en los términos que estamos considerando, aunque su relevancia psicológica dependerá de forma crucial de que dicho análisis se plantee en un nivel adecuado (por ejemplo: la vieja idea de fuerza asociativa puede reinterpretarse en forma de superposición de los espectros correspondientes a dos fenómenos resonantes). Sin embargo, ¿qué forma concreta debe tomar este análisis? Intentaremos responder a esta pregunta mediante la especificación de algunas líneas de estudio que, aun incluyendo algunos elementos intuitivos, pueden ilustrar las afirmaciones que venimos realizando.

Consideremos, por ejemplo, el proceso de solución de problemas en sujetos humanos, y supongamos que se define la solución de un problema como un proceso irreversible que implica una re-estructuración más o menos profunda del campo de conocimiento de un sujeto. En el marco de los modelos de tipo conexionista esto podría traducirse, como veremos inmediatamente, en modificaciones en los estados de equilibrio del sistema de unidades elementales. Supondremos que el funcionamiento de este sistema presenta, a lo largo del proceso de solución (y en cualquier otra situación dada) un conjunto de fluctuaciones de carácter esencialmente aleatorio. En condiciones normales, estas fluctuaciones no se amplificarán y tenderán, en consecuencia, a «regresar» a la situación estable anterior. De acuerdo con los trabajos clásicos de la psicología de la Gestalt (Köhler, 1947; Wertheimer, 1945), consideraremos que la presentación de una situación-problema producirá un aumento en la «tensión» del sistema, cambio que interpretaremos como un alejamiento de la situación de equilibrio, la cual, a su vez, será definida como aquella que presente un nivel global de «energía» mínimo. La energía del sistema es definida, como veremos algo más adelante, en función de la relación entre el estado del sistema cognitivo y su entorno; éste introduce, por definición, una serie de restricciones a las cuales el sistema debe

acomodarse (por ejemplo, las características de la solución o las «reglas del juego» en una situación-problema). Sostendremos también que cuando el nivel de «tensión» supera un determinado umbral (o bien cuando ha transcurrido un período crítico de tiempo) las fluctuaciones aleatorias en el funcionamiento del sistema pueden progresar y ampliarse. En este último caso, el resultado será una bifurcación en el funcionamiento del sistema, que a su vez se traducirá en cambios cualitativos como los que se producen, por ejemplo, en el salto repentino de un dominio de solución a otro (en la terminología de la «teoría de las hipótesis»; Levine, 1975), o en la experiencia subjetiva descrita por los psicólogos de la Gestalt como «comprensión súbita».

La problemática que estamos analizando presenta algunos puntos de contacto con la que se plantea en ciertos ámbitos de la física. El alejamiento de la situación de equilibrio en sistemas donde interactúa un gran número de unidades es producto de un conjunto de ligaduras externas cuyos valores obligan al sistema a alcanzar un estado lejano al de equilibrio. Otra característica básica de este tipo de sistemas, que ha sido incorporada explícitamente por los teóricos conexionistas, es la existencia de ciertos mecanismos de interacción no lineal entre los elementos del sistema. En estas condiciones puede producirse, al contrario que en las estructuras estables, un comportamiento cooperativo de un gran número de unidades que se traduzca en la consecución de un estado estable y de un conjunto de relaciones entre elementos bastante distintas de las que se producían anteriormente.

Algunas líneas más atrás hemos presentado un principio de carácter general que enunciaremos ahora de modo preciso: el sistema de unidades de tipo conexionista mantiene en todo momento un conjunto de valores de activación de unidades y de «pesos» de sus interacciones tal que la energía total del sistema sea la mínima posible, dadas las condiciones del entorno. De esta forma, la consecuencia del aislamiento del sistema sería necesariamente su evolución hasta un estado de equilibrio que, como señalábamos anteriormente, es comparable con su «muerte termodinámica».

Consideremos, por ejemplo, un modelo conexionista que incluya unidades cuyo funcionamiento está asociado a un umbral de activación. De acuerdo con Ackley, Hinton y Sejnowski (1985), la energía correspondiente a una configuración global de activación cualquiera vendría expresada en la siguiente ecuación:

$$E = \sum w_{ij}s_i s_j + \sum u_i s_i$$

donde w_{ij} es la fuerza o «peso» de la conexión entre unidades i y j ; s_i es 1 si la unidad i está en situación de activación, y 0 en caso contrario. u_i es el umbral de activación para la unidad i . Supongamos además que el sistema está ejecutando cualquier tarea (por ejemplo, interpretar una imagen, o solucionar un problema). Será posible, entonces, referirse a la diferencia de energía entre dos configuraciones: la que resultará de aceptar una determinada hipótesis (sobre la interpretación de la imagen, o sobre el camino de solución de un problema), y la que resultaría de no aceptarla. En especial, Hopfield (1982) distingue entre algunas situaciones en las que la evolución más adecuada del sistema es la que le

dirige hacia un mínimo energético local y otros contextos (particularmente ante restricciones importantes procedentes del entorno), en los que será más adecuada la transición hacia un mínimo global. Debe señalarse inmediatamente que los dos tipos de transición que acabamos de mencionar se derivan del análisis de los sistemas gradientes y corresponden a lo que, en el marco de la teoría matemática de la estabilidad estructural, se denomina «regla de Maxwell» (transición hacia el mínimo global) y «regla de retraso» (*delay rule*, transición hacia un mínimo local).

Metropolis, Rosenbluth, Rosenbluth, Teller y Teller (1953) propusieron, en el marco del estudio de las propiedades de sistemas termodinámicos, un algoritmo que posee la propiedad de evitar los mínimos locales de energía para dirigir la transición hacia el mínimo global. Este tipo de planteamiento ha sido aplicado mucho más recientemente al problema de la acomodación a las restricciones del medio (Kirkpatrick, Gelatt y Vecchi, 1983). Ackley, Hinton y Sejnowski (1985) retoman esta idea en una forma adecuada a los sistemas estructurados en paralelo que ejecutan operaciones de procesamiento. En particular, a partir de la diferencia energética entre los estados «activado» o «desactivado» de la unidad k es posible calcular la probabilidad de que dicha unidad supere el umbral de activación. Si una red de unidades sigue una «regla de decisión» de este tipo puede demostrarse que la probabilidad relativa de dos estados energéticos globales sigue la distribución de Boltzmann, que permite calcular la probabilidad de transición entre estados en función de su nivel energético.

La formulación de Hopfield (1982) y de Ackley, Hinton y Sejnowski (1985) corresponde muy exactamente a lo que en el campo de la teoría de la estabilidad estructural es denominado sistema gradiente. En este tipo de formulación matemática los regímenes locales estables son definidos por los atractores estructuralmente estables del campo de dinámicas locales. Si aceptamos algún tipo de principio económico, estos atractores corresponderán a los valores mínimos de una función potencial $V(m;x)$. En general, además de un mínimo global, existirá también un conjunto de mínimos relativos de carácter local. La regla de transición de Maxwell se enunciaría, entonces, en los siguientes términos: dado que diversos atractores C_i , todos ellos estables, están en competición en un punto cualquiera, el atractor C_i hacia el cual tenderá la dinámica del sistema será el de potencial mínimo absoluto. En general, la convención de Maxwell será de aplicación cuando se utilicen valores medios de determinadas magnitudes; en física, por ejemplo, este tipo de regla fue utilizado por Maxwell (a quien obviamente debe su nombre) para describir el cambio «catastrófico» en la densidad cuando un líquido hierve (ecuación de van der Waals). Como señalan Isnard y Zeeman (1976), en contextos más cercanos a las ciencias sociales suele ser de mayor aplicación la «convención de retraso». En cualquier caso, cada campo de aplicación concreto exige la justificación de la regla de transición adoptada.

La formulación de Ackley, Hinton y Sejnowski (1985) toma explícitamente el nombre de «máquina de Boltzmann». No cabe duda de que se trata de una denominación muy adecuada y sugerente, aunque represente únicamente un tipo concreto de modelo conexionista. En particular, es necesario señalar que en una red con unidades de funcionamiento completamente continuo el mecanismo de

minimización de la aportación energética de una unidad consistiría en aumentar o disminuir el estado de activación de forma proporcional al input recibido. Al margen de estas cuestiones (que, no obstante, están lejos de ser irrelevantes), la formulación que estamos considerando hereda algunos de los problemas del enfoque boltzmanniano y, muy particularmente, su carácter «excesivamente regular» o continuo. El problema es el mismo que se plantea a la termodinámica clásica, ligado esencialmente al carácter necesariamente creciente de la magnitud entropía. A partir del enfoque estadístico de Boltzmann, la entropía contendrá el «grado de desinformación» del observador macroscópico y representará el número de configuraciones microscópicas compatibles con cierto estado de equilibrio. La entropía de Boltzmann asocia al sistema una idea de orden, de manera que el segundo principio de la termodinámica clásica se convierte en una ley de desorganización progresiva. Abundando en algunas ideas ya mencionadas, afirmamos explícitamente que la formulación de un sistema distribucional interactivo para los procesos cognitivos en términos meramente boltzmannianos no puede dar cuenta de todo un conjunto de cambios cualitativos que, a nuestro entender, forman parte inextricablemente de los procesos psicológicos. De ahí, por una parte, la formulación del concepto de estructura disipativa y, por otra, la necesidad, que nosotros afirmamos, de recurrir a una visión semejante a la de la teoría de la estabilidad estructural para la modelización de la dinámica cualitativa del sistema cognitivo.

Retomaremos en este punto la cuestión de posibles fenómenos resonantes en el sistema cognitivo. Es bien sabido que la metáfora de la resonancia ha jugado un papel de cierta importancia en el ámbito de la psicología, aunque casi siempre de forma un tanto inconcreta y con un escaso nivel de desarrollo. Por ejemplo Gibson (1966) utiliza el término «resonancia» para referirse a los supuestos mecanismos internos que constituyen la percepción. Este tipo de argumentos ha sido recogido, por ejemplo, por Shepard (1984) quien, en su análisis de las representaciones internas, muestra que la metáfora de la resonancia permite incluir en los modelos de la representación ciertas propiedades que pueden ser caracterizadas como «restricciones ecológicas» introducidas a lo largo de la filogénesis. Aunque Gibson se refiere básicamente a lo que él denomina «resonancia de información», pueden encontrarse también análisis en un nivel más «energético» como, por ejemplo, las importantes referencias, en el ámbito de la fisiología, a fenómenos de «resonancia neural». Pueden citarse, a este respecto, las ideas de patrones de interferencia (Lashley, 1942), circuitos reverberantes (Ashby, 1954; Rashevsky, 1948) o de asambleas celulares y secuencias de fase reverberantes (Hebb, 1949). Más recientemente, el concepto de resonancia neural ha sido utilizado tanto en análisis específicos relacionados, por ejemplo, con el funcionamiento de los receptores sensoriales, como en construcciones teóricas de gran envergadura como el sistema propuesto por Grossberg (1980).

En términos más generales, es posible demostrar, desde un análisis puramente matemático, la idea fundamental de que *la resonancia es una propiedad que emerge de forma natural del funcionamiento de cierto tipo de redes*. Greene (1962a, 1962b) muestra que si la información es representada en forma de señales *graduadas*, entonces incluso una red compuesta solamente por elementos li-

neales poseerá resonancias características. Este tipo de sistema presentará dos propiedades fundamentales: a) las «resonancias» que representen información significativa tenderán a ser estables frente a perturbaciones aleatorias; b) configuraciones informativas extremadamente complejas pueden ser representadas a través de un pequeño número de intensidades (es decir, de niveles de activación de unidades elementales o conjuntos de tales unidades). Resulta evidente que estas propiedades son altamente deseables en el funcionamiento del sistema cognitivo, y también lo es su relevancia para una visión «dinámica» del funcionamiento de dicho sistema, en la línea propuesta por la orientación conexionista.

Es interesante analizar con mayor detenimiento algunas de las potencialidades de la metáfora de resonancia. Se ha indicado anteriormente que el análisis de los fenómenos resonantes es posible desde una perspectiva topológica. Como señala Thom (1972), generalmente la noción de resonancia no es analizada matemáticamente de forma explícita más que en el caso muy particular de un oscilador lineal sometido a un impulso de frecuencia igual a su frecuencia propia. Sin embargo, incluso en este caso concreto un análisis sencillo de los estados de equilibrio del sistema y de su dinámica general (descrita por una ecuación relativamente simple) revela la existencia de regiones en las cuales la función que describe el sistema puede tomar hasta tres valores (indicando, de este modo, la presencia de bifurcaciones). Resulta evidente, entonces, que incluso los sistemas resonantes más sencillos muestran un comportamiento en el cual existe la posibilidad de cambios cualitativos que, a nuestro entender, deben ser analizados y formar parte de cualquier modelo de carácter mínimamente general.

Complicando algo más la cuestión, supongamos la existencia de dos diapasones D y D' , e imaginemos que D está vibrando. Si aproximamos D a D' , entonces D' comenzará a vibrar en resonancia con D , de forma que una parte de la energía cinética de D se transferirá a D' . El producto de los osciladores lineales puede ser analizado en términos topológicos (Thom, 1972, 1980) tomando como ejemplo de sistema dinámico el de dos círculos sobre cada uno de los cuales se da cierto campo de vectores constante. Puede demostrarse que el sistema resultante no es puramente el producto topológico de los sistemas originales y que, en mayor o menor medida, existirá una degeneración hacia un régimen dinámico común para todo el sistema, el régimen de resonancia. En ciertos casos (matemáticamente, cuando la pendiente de las rectas que definen las trayectorias del campo resultante del producto de los dos sistemas originales es un número irracional), el sistema global presentará resonancia difusa, bastante inestable y fluctuante. En otros casos (cuando la pendiente de las rectas es representada por un número racional), cada elemento perderá totalmente su independencia inicial y ambos entrarán en una dinámica única, el sistema resonante. Expresando en otros términos la idea básica del análisis de sistemas resonantes, dos sistemas dinámicos sólo pueden intercambiar energía por resonancia si presentan modos vibratorios que posean ciertas características comunes que admiten un análisis basado en conceptos topológicos. Esta idea se deriva, en parte, de la mecánica hamiltoniana, que reconduce el estudio de los movimientos de ciertos sistemas de puntos al estudio geométrico del correspondiente espacio de fases. En particular, puede afirmarse que el sistema oscilante posee una interpretación morfológica en términos

de las singularidades de una hamiltoniana (Thom, 1972). Para una energía global E , puede admitirse que las características vibratorias de un sistema S están enteramente definidas por una figura $T(e)$ (función de E) en un cierto espacio euclidiano: esta figura puede ser llamada espectro del sistema. Cuando dos sistemas, S y S' , entran en interacción, existe una identificación entre los dos espacios asociados a ellos; en la medida en que los dos espectros puedan ser superpuestos existirá la posibilidad de interacción resonante entre ellos. La resonancia será aguda si, para un conjunto de pares de valores $E-E'$, los dos espectros pueden estar exactamente superpuestos.

Pensamos que este tipo de análisis posee implicaciones importantes para la temática que estamos tratando. En efecto, si puede demostrarse que los sistemas dinámicos del tipo propuesto en los modelos distribucionales incorporan como característica global la presencia de dinámicas resonantes, y si este tipo de dinámicas es analizable en los términos que acabamos de exponer (sin duda de modo muy simplificado y esquemático), entonces es inmediatamente evidente la factibilidad de una descripción en términos topológicos de las similitudes cualitativas entre patrones de activación y, además, se refuerza claramente una concepción global basada en la suposición de que el sistema cognitivo humano funciona en gran medida a partir de procesos de tipo continuo que incorporan algunas de las propiedades que estamos considerando. En términos más particulares, el desarrollo sistemático de la metáfora de la resonancia reforzaría la posibilidad de plantear un enfoque dinámico de los procesos cognitivos.

El enfoque de los procesos cognitivos en términos de estabilidad estructural permite conciliar dos hechos superficialmente dispares: por un lado, la dinámica propia de un sistema basado en conexiones continuas, en el cual la información es representada en forma de niveles de activación de unidades elementales; por otra parte, la presencia de discontinuidades significativas en la conducta del sujeto. Estas discontinuidades pueden abarcar desde fenómenos tipo umbral, básicos en psicofísica e incluidos como elementos hipotéticos en muchos modelos distribucionales interactivos, hasta cambios cualitativos de tipo mucho más global como la aparición de discontinuidades en el aprendizaje (Kolers y Duhnicky, 1985), de estadios de desarrollo cualitativamente distintos en sistemas evolutivos (Molenaar, 1986; Molenaar y Openheimer, 1985), de transformaciones repentinas en la solución de problemas, como proponía clásicamente la psicología de la Gestalt, de fenómenos todo-nada en el ámbito del aprendizaje, etc. Una teoría basada en el concepto de estabilidad estructural reintroduce la posibilidad de incluir este tipo de fenómenos en el marco general de la psicología científica, de la cual han estado en algunos casos parcialmente excluidos por su carácter vago y poco definido. Parece claro, a partir de las ideas que hemos expuesto, que esta exclusión debe ser atribuida esencialmente a la falta de incorporación a la psicología, hasta este momento, de instrumentos de modelización que permitieran diseñar una aproximación científica a este tipo de fenómenos. El desarrollo de modelos matemáticos basados en la noción de estabilidad estructural puede contribuir a llenar de forma decisiva este vacío.

REFERENCIAS

- Ackley, D.W., Hinton, G.E. & Sejnowski, T.J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9, 147-169.
- Anderson, J.R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Ashby, W.R. (1954). *Design for a brain*. New York: Wiley.
- Broadbent, D. (1985). A question of levels: Comment on McClelland and Rumelhart. *Journal of Experimental Psychology: General*, 114, 189-192.
- Cohen, J.D., Dunbar, K. & McClelland, J.L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332-361.
- Fahlman, S.E. (1979). *NETL: A system for representing and using real-world knowledge*. Cambridge: MIT Press.
- Fodor, J. (1975). *The language of thought*. New York: Crowell.
- Fodor, J. & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Fodor, J. & McLaughlin, B.P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, 35, 183-204.
- Gibson, J.J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Greene, P.H. (1962a). On looking for neural networks and cell assemblies that underlie behavior. I. A mathematical model. *The Bulletin of Mathematical Biophysics*, 24, 247-275.
- Greene, P.H. (1962b). On looking for neural networks and cell assemblies that underlie behavior. II. Neural realization of the mathematical model. *The Bulletin of Mathematical Biophysics*, 24, 395-411.
- Grossberg, S. (1980). How does the brain build a cognitive code? *Psychological Review*, 87, 1-51.
- Hebb, D.O. (1949). *The organization of behavior*. New York: Wiley.
- Hinton, G.E. & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74-95.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*.
- Isnard, C.A. & Zeeman, E.C. (1976). Some models from catastrophe theory in the social sciences. En L. Collins (Ed.): *The use of models in the social sciences*. London: Tavistock.
- Jacoby, L.L. (1983). Remembering the data: Analyzing interaction processes in reading. *Journal of Verbal Learning and Verbal Behavior*, 22, 485-508.
- Johnson-Laird, P.N. & Wason, P.C. (1977). *Thinking. Readings in cognitive science*. Cambridge: Cambridge University Press.
- Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, 220, 671-680.
- Köhler, W. (1947). *Gestalt Psychology*. New York: Liveright.
- Kolers, P.A. & Ducknicky, R.L. (1985). Discontinuity in cognitive skill. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 655-674.
- Lashley, K.S. (1942). The problem of cerebral organization in vision. In H. Kluver: *Biological Symposia, VII*. Lancaster: Jaques Cattell Press.
- Levine, M. (1975). *A cognitive theory of learning*. Hillsdale: Erlbaum.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- McClelland, J.L. & Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159-188.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953). Equation of state calculations for fast computing machines. *Journal of Chemical Physics*, 6, 1087.
- Molenaar, P. (1986). Issues with a rule sampling theory of conservation learning from a structuralist point of view. *Human Development*, 29, 137-144.
- Molenaar, P. & Oppenheimer, L. (1985). Dynamic models of development and the mechanistic-organismic controversy. *New Ideas in Psychology*, 3, 233-242.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, 76, 165-178.
- Morton, J. (1979). Facilitation of word recognition: Experiments causing change in the logogen model. In P.A. Kolers, M.E. Wrostad & H. Bowna: *Processing of visible language*. New York: Plenum.
- Morton, J. (1981). The status of information processing models of language. *Philosophical transactions of the Royal Society of London*, 295, 387-395.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4, 135-183.
- Prigogine, I. (1983). *¿Tan sólo una ilusión?* Barcelona: Tusquets.
- Pylyshyn, Z.W. (1986). *Computation and cognition*. Cambridge: MIT Press.
- Rashevsky, N. (1948). *Mathematical Biophysics*. Chicago: University of Chicago Press.

- Rivière, Á. (1984). *Sobre la multiplicidad de las representaciones. Un viaje por los vericuetos de los lenguajes del pensamiento*. Symposium sobre Actividad Humana y Procesos Cognitivos. Madrid.
- Rivière, Á. (1986). *Razonamiento y representación*. Madrid: Siglo XXI.
- Rumelhart, D.E. & McClelland, J.L. (1985). Levels indeed! A response to Broadbent. *Journal of Experimental Psychology: General*, 114, 193-197.
- Rumelhart, D.E. & McClelland, J.L. (1986). *Parallel distributed processing. Explorations in the microstructure of cognition*. Cambridge (Mass.): MIT Press.
- Seidenberg, M.S. & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Shepard, R.N. (1984). Ecological constraints on internal representation. Resonant kinematics of perceiving, imagining, thinking and dreaming. *Psychological Review*, 91, 417-447.
- Smolensky, P. (1988a). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1-23.
- Smolensky, P. (1988b). Connectionism, constituency and the language of thought. In Loewer, B. & Rey, G. (Eds.), *Meaning in mind: Fodor and his critics*. Oxford: Blackwell.
- Smolensky, P. (1989). Connectionism and constituent structure. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié & L. Steels (Eds.), *Connectionism in perspective*. North-Holland: Elsevier Science Publishers.
- Thom, R. (1972). *Stabilité structurelle et morphogénèse*. Reading: Benjamin.
- Thom, R. (1980). *Modèles mathématiques de la morphogénèse*. Paris: Bourgois.
- Van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, 14, 355-384.
- Van Orden, G.C., Pennington, B.F. & Stone, G.O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, 97, 488-522.
- Vega, M. de (1984). *Introducción a la psicología cognitiva*. Madrid: Alianza.
- Wertheimer, M. (1945). *Productive thinking*. New York: Harper and Row.

