Capítulo 11

MODELOS COMPUTACIONALES EN NEUROPSICOLOGÍA COGNOSCITIVA

A lo largo de la historia, los "investigadores de la mente", por llamar de algún modo a todos aquellos que han dedicado gran parte de su vida a desentrañar los misterios de esa peculiar "herramienta" con que cuenta el hombre, se han valido de analogías para intentar explicar la naturaleza de las diversas funciones psicológicas. Desde Aristóteles, con su idea de que la memoria era una suerte de "tablilla de cera", pasando por Hume y sus pensamientos encadenados "como vagones de un tren", hasta Freud, que usa su modelo hidráulico para explicar el flujo de la libido en la vida mental del hombre, las metáforas han sido ampliamente utilizadas como útiles y ricos recursos teóricos para acercarnos a la verdadera naturaleza de "lo mental" (Rumelhart, 1989).

Sin embargo, desde hace unos treinta años, se ha venido imponiendo un nuevo paradigma en la explicación psicológica basado en la analogía entre la mente y el computador. Esta nueva tendencia ha sido llamada por muchos teóricos "el enfoque del procesamiento de información", y es la psicología cognitiva la ciencia que se ha encargado de estudiar la mente desde esta nueva perspectiva. (Eysenck y Keane, 1995). Pero, ¿de dónde surge la idea de que el computador -y en particular, los procesos computacionales-, pueden ayudarnos a develar el misterio que oscurece la verdadera naturaleza de los procesos cognitivos? El origen de esta nueva concepción es difícil de rastrear, sin embargo, es posible rescatar dos importantes hechos que aceleraron la proliferación y posterior aceptación de los modelos computacionales al interior de la psicología. Por un lado, hacia 1950, muchos teóricos empezaron a cuestionar seriamente la viabilidad del modelo conductista de explicación psicológica. La idea de atender únicamente a los estímulos y respuestas de un organismo para

explicar su comportamiento, generaba importantes dudas que no parecían tener respuesta dentro de una teoría que trataba al cerebro y a sus funciones como una "caja negra". Karl Lashley fue uno de los primeros pensadores en poner en duda la capacidad explicativa del conductismo, pues creía que ciertos procesos psicológicos, como la producción espontánea de lenguaje, involucraban mecanismos detallados al interior del aparato cognitivo del individuo que no podían ser explicados únicamente a partir de cadenas de estímulos y respuestas. De modo similar, Noam Chomsky probaría que el modelo conductista no daba cuenta del complicado aparato sintáctico con que habría de contar el hombre para producir oraciones en un lenguaje (Johnson-Laird, 1988, 23). Estas y otras críticas generaron, entonces, la necesidad de estudiar los procesos psicológicos desde otra perspectiva.

El otro evento crucial en el origen del paradigma del procesamiento de información fue la creación de los primeros computadores digitales capaces de almacenar sus propios programas, a cargo del matemático John von Newmann. Él, sin saberlo, compartía la idea de otro matemático –el inglés Alan Turing-, quien había diseñado un magnífico recurso conceptual, denominado "máquina de Turing", por medio del cual podían descomponerse casi todo tipo de procesos matemáticos (que, según se creía, necesitaban siempre de la intervención del pensamiento) en una cadena de operaciones básicas totalmente mecánicas. (Turing, 1937). Basándose en su invención, Turing conjeturó que los procesos mentales podrían operar de modo similar a sus máquinas, es decir, algorítmicamente. La conjunción de estas ideas motivó a muchos investigadores a construir no sólo modelos computacionales que simularan operaciones intelectuales, sino a descubrir los "algoritmos" que subyacen a los procesos cognitivos humanos. Dentro del primer grupo se encontraban Allen Newell y Herbert A. Simon, quienes, a mediados de los 50, construyeron un computador que por primera vez simulaba procesos inteligentes, como toma de decisiones y resolución de problemas. Su *General Problem Solver* contribuiría a iniciar, entre otras cosas, una importante rama de investigación, conocida como *Inteligencia Artificial*, que busca crear aparatos que simulen procesos inteligentes a partir de operaciones mecánicas-computacionales sobre símbolos. (Newell y Simon, 1972; Newell y Simon, 1976; Boden, 1977).

Al segundo grupo, en cambio, pertenecen los psicólogos cognitivos que, valiéndose del bagaje conceptual y de las posibilidades prácticas de la ciencia computacional, han desarrollado diversos tipos de modelos expli-

cativos de la mente, sobre el supuesto de que el nivel funcional o algorítmico de la cognición es, efectivamente, un buen nivel de explicación. En realidad, fue la aceptación de esta idea la que en últimas habría marcado el origen de la psicología cognitiva. Hasta hace relativamente poco tiempo, quizás sólo hasta principios del siglo XX, la investigación en psicología transitaba por tres caminos distintos: el tradicional enfoque dualista, el difícil conductismo y el radical materialismo. La imposibilidad de explicar el funcionamiento del alma llevó al dualismo a su inminente fracaso, la inexactitud de sus explicaciones y predicciones arruinaron el proyecto conductista en psicología, y la profunda ignorancia en los procesos físicos del cerebro impiden, por lo pronto, que la sola neurociencia pueda ser considerada como una teoría psicológica adecuada. No obstante, la opción computacional en psicología no se ha visto libre de críticas. Desde el momento mismo en que el término "ciencia cognitiva" fue introducido en el ámbito de la investigación psicológica -probablemente en 1956, durante una reunión en el MIT a la que asistieron varios de los forjadores de esta nueva postura-, muchos se han pronunciado en contra de ella. Debe reconocerse, en todo caso, que a partir de los 70, la mayor parte de los psicólogos teóricos aceptan un buen número de los postulados de la ciencia cognitiva, y cada vez más se hace común entre ellos el uso de términos heredados del modelo de procesamiento de información.

En consecuencia, se puede definir de un modo muy general a la ciencia cognitiva como una disciplina en la que confluyen diversas vertientes de investigación (como filosofía, neurociencia, inteligencia artificial y lingüística, entre otras), y que tiene por objeto explicar científicamente el modo en que el cerebro efectúa toda clase de funciones cognitivas, creando así mapas conceptuales que aclaren la naturaleza de los procesos neuronales que subyacen a las manifestaciones comportamentales de los seres humanos. (Eysenck y Keane, 1995, 4; Parkin, 1996; Seifert y Shafto, 1994, 410). Algunos teóricos, para resumir esta definición, consideran que su objetivo es simplemente descubrir "la arquitectura (funcional) de la mente".

Con el tiempo, la psicología cognitiva, heredera de las ideas de la ciencia cognitiva, se ha dividido en al menos tres grandes ramas: aparecen, en primer lugar, los psicólogos cognitivos experimentales que se resisten a aceptar totalmente los postulados del modelo del procesamiento de información, aunque apliquen ciertas nociones de la ciencia cognitiva "más tradicionales". (Eysenck y Keane, 1995, 3). También están, en segun-

do lugar, los partidarios de la neurociencia cognitiva, quienes construyen complejos modelos computacionales para explicar la manera en que surgen las manifestaciones comportamentales en las personas a partir de su actividad cerebral. (Rugg, 1997). Y, finalmente, se encuentran los neuropsicólogos cognitivos "quienes afirman que la investigación de los patrones de impedimentos cognitivos en pacientes con daño cerebral puede proveer[nos] de valiosa información concerniente a la cognición humana normal". (Cfr. Eysenck y Keane, 1995, 3. Ver también en: Shallice, 1991, 429). Como hija de la ciencia cognitiva, la neuropsicología cognitiva (o cognoscitiva, para usar el término castizo) se ha valido también de la noción de procesamiento de información y, con ella, de la construcción de modelos computacionales para formular sus teorías sobre el funcionamiento normal de la cognición humana. En lo que sigue se analizarán los dos principales recursos teóricos computacionales adoptados por esta ciencia: el modelo modular y el conexionista.

11.1 Del computador al cerebro: modelos modulares

Uno de los más importantes supuestos de la neuropsicología cognoscitiva moderna es el de que diferentes funciones mentales son llevadas a cabo por distintas regiones cerebrales. Esta idea es conocida como *especificidad neurológica* y se ha venido validando, a través del tiempo, gracias al estudio de pacientes que demuestran incapacidades específicas en virtud de daños cerebrales muy localizados. De esta forma, las investigaciones llevaron a los neuropsicólogos a formular modelos en los que tales funciones operasen por separado, es decir, con independencia de las demás. Por ello se empezó a pensar en el hecho de que los modelos deberían permitir la ocurrencia de lo que se ha denominado *disociación doble clásica*, una característica del cerebro que consiste en que el mal funcionamiento de determinada operación cognitiva no trastorne el normal desarrollo de otra. (Parkin, 1996). Y junto con el hecho de que la mayor parte de los desórdenes registrados en neuropsicología clínica son altamente específicos y cualitativamente diversos, algunos teóricos empezaron a tratar las distintas funciones cognitivas como subsistemas (Posner, 1978) o módulos (Marr, 1982; Fodor, 1986) de un complejo funcional organizado (Shallice, 1991). De hecho, la idea de que ciertas funciones estén claramente delimitadas y operen con relativa independencia de las demás, pare-

ce haber estado a la base de viejas concepciones de la neuropsicología clínica. Tal es el caso de la *alexia pura* descrita por Déjèrine en 1882 (Cfr. capítulo 5). Él, no sin razón, calificó de 'pura' a una cierta dificultad muy específica en la comprensión del lenguaje escrito, en contraposición a las *alexias impuras* que referían a daños en más de una función lingüística. Probablemente si Déjèrine no hubiera presupuesto un cierto tipo de individualidad (o fraccionamiento) funcional en el cerebro, su clasificación no tendría mayor sustento.

Sin embargo, sólo hasta comienzos de la década de los 80, los principios de la especificidad neurológica y de la disociación funcional se convirtieron en verdaderos axiomas para una teoría sobre el funcionamiento de la mente. Quizás la más estructurada exposición de una teoría como ésta fue la formulada por Fodor en 1983 -luego de un primer intento desarrollado por Marr (1976, 1982)-, en lo que él mismo llamó: la modularidad de la mente. Además, esta nueva teoría tiene una doble importancia en psicología cognitiva: por un lado, constituye la primera aproximación teórica seria y suficientemente amplia que se vale de recursos de la ciencia computacional para explicar los mecanismos que subyacen a nuestras operaciones mentales. Tan es así, que el propio término 'modularidad' es heredado del lenguaje de los programadores de computadores, y refiere al principio "de que es importante hacer a los diferentes componentes del programa [de computador] tan independientes uno del otro como sea posible". (Parkin, 1996). Y, por otro lado, resulta ser el primer intento de formulación de una arquitectura computacional de la cognición humana, esto es, una propuesta para entender la mente desde un punto de vista funcional y no físico, en donde se enfatiza en la cercanía operativa entre el programa del computador y la mente, y no en la similaridad estructural entre el cerebro y el *hardware* del aparato.

Con esto en mente, Fodor propone una serie de características que deben tener los diferentes sistemas cognitivos para que puedan ser considerados módulos. Cabe aclarar en este punto que la arquitectura funcional por él propuesta distingue tres sistemas operativos distintos: los transductores, los sistemas de entrada y los sistemas centrales, de los cuales sólo los segundos presentan la estructura modular que ahora se examinará. (Fodor, 1986, 69). En términos generales, un transductor puede entenderse simplemente como un órgano sensorial mono-modal, es decir, sensible a un tipo cualitativamente específico de estímulo proximal, que provee de la materia prima a los sistemas de

entrada para que éstos tengan datos perceptuales sobre los cuales efectuar sus computaciones. Los sistemas centrales, por el contrario, parecen ostentar una estructura no-modular y se entienden como los encargados de realizar operaciones cognitivas muy complejas, como fijación de creencias y toma de decisiones. Dada la cantidad de operaciones implicadas en este tipo de procesos cognitivos superiores y la gran diversidad de contenidos sobre los cuales operan, su naturaleza sigue siendo bastante desconocida y, valga decir, no sólo en el ámbito de los modelos modulares, este asunto es también una piedra de toque en la mayor parte de las teorías de la mente. Por ello, la exposición se limitará a la naturaleza de los sistemas de entrada, cuya función puede definirse, en términos muy generales, así:

> Los sistemas de entrada operan con el fin de suministrar información a los procesadores centrales; en concreto, sirven de mediadores entre los productos de los transductores y los mecanismos cognitivos centrales efectuando la codificación de las representaciones mentales que constituyen el ámbito de operación de tales mecanismos centrales. (Fodor, 1986, 69)

Pero esta mediación no debe entenderse como una 'traducción' sin más, en la medida en que los sistemas de entrada están pensados para preservar el contenido informativo, que no siempre es respetado en una traducción. De este modo puede considerarse a los sistemas de entrada como una especie de 'sistemas de inferencia', cuyas 'premisas' "son las representaciones transducidas de las configuraciones estimulares proximales, y [que tienen] como 'conclusiones' las representaciones de la naturaleza y la distribución de los objetos distales". (Fodor, 1986, 70). Lo cual equivale a decir, en términos más sencillos, que los transductores le brindan a los sistemas de entrada los datos perceptuales 'crudos', en virtud de los cuales estos últimos pueden determinar la distribución de los objetos en la realidad.

Ahora bien, a nivel de su definición, los sistemas de entrada pueden caracterizarse, en términos generales, como "sistemas cognitivos modulares [que] son específicos de dominio, fijados de modo innato, compactos, autónomos y no ensamblados". (Fodor, 1986, 63). Pero adentrándonos un poco más en la teoría, se puede decir que tales sistemas tienen ciertas propiedades específicas, a saber:

1. Aún no conocemos el número determinado de sistemas de entrada. Aunque en principio parecería plausible suponer que existen sólo seis tipos -cinco sistemas perceptuales (uno para cada sentido) y uno lingüístico-, la evidencia nos demuestra que los procesos perceptuales pueden descomponerse a su vez en varios subprocesos que operan con independencia de los demás. Así, por ejemplo, a nivel visual la experiencia nos muestra que la identificación de la forma, del color o de la tridimensionalidad de los objetos, son procesos que ocurren por separado. De igual modo, se ha llegado a conclusiones similares en cuanto a la comprensión de las oraciones escuchadas: en estos casos intervienen sistemas más específicos que se encargan, v.gr. de asignar descripciones gramaticales a los enunciados verbales percibidos o de reconocer la familiaridad de la voz del interlocutor, entre otros.

2. Los sistemas de entrada son específicos de dominio, lo cual significa que cada módulo está habilitado para procesar un único tipo específico de *input* o dato de entrada. Así, un módulo encargado de, por ejemplo, el reconocimiento de rostros, no puede recibir ni procesar información relativa a la visión de objetos de uso doméstico.

3. El funcionamiento de los sistemas de entrada es obligatorio, es decir, demuestra un carácter 'mandatario'. Esto significa que los sistemas de entrada están obligados a operar con la información que reciben, siempre y cuando puedan hacerlo. Cierta evidencia (Fodor, 1986, 83) nos demuestra, por ejemplo, que al solicitarle a algunos sujetos normales que atendieran únicamente a las propiedades acústico-fonológicas de una oración, eran incapaces de dejar de identificar las palabras que la componían. En consecuencia, mientras los sistemas de entrada permanezcan intactos, no podemos ejercer control selectivo alguno sobre su funcionamiento, y ellos estarán obligados a recodificar la información recibida de los transductores para enviarla a los sistemas centrales.

4. Aunque el procesamiento perceptivo, que va "de abajo hacia arriba" (desde el producto más inmediato del transductor hasta el más lejano del sistema de entrada), es obligatorio e "inaccesible a la conciencia", el acceso a los productos de dicha operación por parte de los sistemas centrales, que se efectúa "de arriba hacia abajo", es más restringido. De esta forma, cuanto más alejada esté la información del transductor, cuanto más elaborada sea, más fácil es el acceso de los sistemas centrales a ella.

5. Los sistemas de entrada realizan sus operaciones con increíble rapidez, al punto en que puede resultar difícil distinguir si la celeridad con que se logra el reconocimiento de, por ejemplo, una oración en castellano, es debida a la velocidad de uno u otro módulo involucrado en el proceso.

6. Los sistemas de entrada se hallan informativamente encapsulados, lo cual quiere decir que los módulos efectúan sus operaciones en aislamiento de lo que ocurre en otros lados. Así, en el caso del reconocimiento de un enunciado verbal, se hace necesaria la participación de varios módulos: de análisis fonológico, fonético, léxico, sintáctico, etc. No obstante, aún cuando el flujo de información siga una secuencia ordenada, de forma que a un módulo intermedio sólo podrá tener acceso el inmediatamente anterior y, a su vez, éste sólo podrá acceder al inmediatamente posterior, cada uno de ellos es autónomo y no tiene poder sobre las computaciones de ningún otro, aparte de él mismo. En este sentido, un módulo como el de reconocimiento de fonemas nunca podrá reemplazar las funciones del módulo de análisis sintáctico, y viceversa. La capacidad de un sistema cognitivo para participar en el funcionamiento de otro es lo que se ha llamado *penetración cognitiva* y, valga aclarar, aunque tal característica no se da a nivel de los módulos, sí parece ocurrir a nivel de los sistemas centrales.

7. Los sistemas de entrada se hallan asociados a configuraciones neuronales fijas, establecidas según procederes peculiares de la ontogénesis y, además, presentan pautas de deterioro características y específicas para cada uno.

Todas estas características, por un lado, resumen la naturaleza de algunos modelos de sistemas aislados aceptados durante mucho tiempo y, por el otro, ofrecen un buen número de condiciones para las posteriores estrategias de construcción de modelos cognitivos. Debe tenerse en cuenta, en todo caso, que la aceptación sin más de tales propiedades, no constituye una buena estrategia para la formulación de modelos computacionales de la mente, ya que para hacerlo deben aceptarse –como lo han señalado algunos teóricos- por lo menos los siguientes cuatro supuestos:

1. Los sistemas cognitivos son cualitativamente similares en los humanos. (Eysenck y Keane, 1995, 14).

2. Existe una correspondencia relativamente *isomórfica* entre la estructura de la mente y la del cerebro.

3. La teoría modular es *transparente*, es decir, debe aceptarse que "la ejecución alterada del paciente da las pautas necesarias para saber qué componente modular del sistema ha quedado interrumpido". (Manning, 1992, 77).

4. La aceptación de la teoría de la modularidad implica el supuesto de *sustractividad*, "por el que se entiende que la ejecución de un paciente con lesión cerebral refleja el procesamiento cognitivo total menos el de aquellos sistemas que han sido dañados". (Manning, 1992, 77).

En consecuencia, la aceptación de estos supuestos y de las características anteriormente mencionadas al respecto de los módulos, hicieron de la teoría modular de la mente un buen modelo explicativo en neuropsicología. Y así, la idea de que a partir del análisis de pacientes con alteraciones específicas se pudieran establecer el o los módulos afectados, motivó la formulación de diversos diagramas que representaban el desempeño normal de las funciones cognitivas.

11.1.1 Memoria y modularidad

Como ya se ha anotado, los modelos modulares en neuropsicología cognoscitiva están basados en la arquitectura funcional de los computadores digitales, particularmente de aquellos que ostentan un procesamiento lineal secuencial heredado de los prototipos de von Newmann. Aunque los adelantos tecnológicos han hecho que las características de cada uno cambien en detalle, sus componentes principales –siguiendo a Johnson-Laird (1988, 144)- aún se mantienen en la mayoría de casos. En términos generales, dichos componentes son: un procesador central, encargado de controlar las acciones del aparato; un sistema de memoria en el que se almacenan datos, programas y resultados, y un sistema de entrada-salida, que comunica al procesador central con los sistemas periféricos (teclado, scanner, ratón, impresora, etc.). Pero, ¿qué características de estos procesadores se ven reflejadas en los modelos modulares? Para contestar esta pregunta, estudiemos brevemente y de forma muy abstracta el funcionamiento de un computador secuencial, enfatizando en las propiedades de su sistema de memoria (Fig. 11-1).

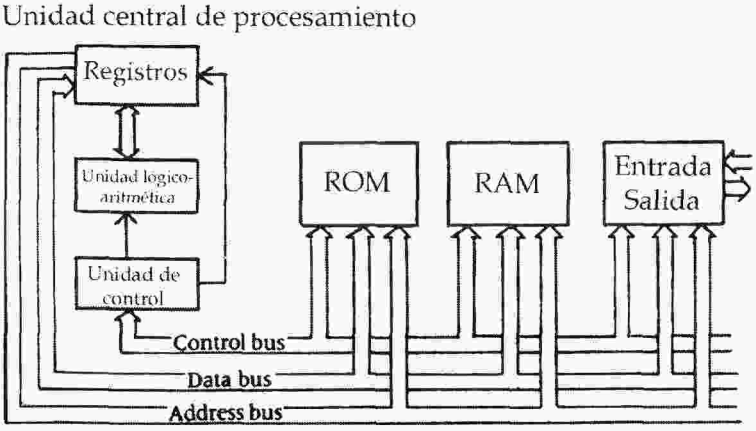

Fig. 11-1
Diagrama simplificado de la arquitectura del computador (Modificado de Johnson-Laird, 1988, p.145)

Una vez han entrado, a través de un sistema periférico de *input* cualquiera, ciertos datos de entrada, la unidad de procesamiento central opera con ellos en virtud de los programas que tenga almacenados en memoria. El proceso, por supuesto, es mucho más detallado, pero baste con decir aquí que este sistema de memoria está dividido en tres subsistemas distintos, cada uno de los cuales le permite acceso al sistema central de acuerdo con una localización especificada en código binario. Distinguimos, pues, un sistema de memoria permanente (ROM), cuyos datos sólo pueden ser leídos y, normalmente, permanecen inalterados. Un sistema de memoria de acceso (RAM), cuyos contenidos pueden ser modificados por el sistema central y que guarda los datos de los programas y los resultados de las computaciones efectuadas por ellos. Y un último "sistema" de memoria secundario, compuesto por los discos magnéticos o disquetes, cuyo uso es temporal y no depende del aparato mismo.

Ahora bien, se ha hablado ya (Cfr. capítulo 7) de uno de los más importantes modelos modulares de memoria, el propuesto por Atkinson y Shiffrin en 1968. En éste se decía que la estimulación externa entraba primero a un almacén sensorial donde, después de un proceso de codificación inicial, pasaba a un segundo módulo conocido como *memoria a corto plazo*. Esta información podría luego pasar a un tercer almacén, la *memoria a*

largo plazo, o ser simplemente descartada. Además, atendiendo al hecho de que parte fundamental de la memoria humana es su capacidad de recobrar información, el modelo planteaba una estructura bidireccional entre la memoria a corto plazo y la memoria a largo plazo. (Parkin, 1995, 4). El modelo, si bien tiene las características de uno modular, discrepa en varios aspectos del funcionamiento del computador, debido al hecho de que no está inspirado en él (en realidad, son casi contemporáneos). Tales diferencias, siguiendo a Johnson-Laird, pueden resumirse del siguiente modo:

> Los computadores tienen una memoria principal (RAM) que se ubica a medio camino entre el almacenamiento permanente a largo plazo (en cinta o disco) y el transitorio registro a corto plazo de los resultados intermedios de las computaciones (en registros). Las personas, sin embargo, parecen estar equipadas con una memoria a largo plazo, que es de fácil acceso (como RAM) aunque relativamente permanente (como la cinta o el disco). Ellas también tienen recuerdos sensoriales, mientras que los computadores, que para la mayor parte carecen de órganos sensibles, no hay tales recuerdos. (Johnson-Laird, 1988, 151).

No obstante, algunos años más tarde (en 1977), Baddeley y Hitch complementarían este modelo con sus investigaciones sobre la memoria a corto plazo, acercando notablemente el funcionamiento de la memoria humana a la del computador (Fig. 11-2). El resultado fue un modelo cuasimodular que combinaba las características descritas por Fodor en tres subsistemas generales (almacén sensorial, memoria a corto plazo y memoria a largo plazo), con otros dos subsistemas subsidiarios, el sistema de registro viso-espacial y el bucle (*loop*) articulatorio, que ostentaban las mismas características de los módulos salvo su especificidad de dominio –por el hecho de ser polimodales. (Baddeley, 1995). Así, el sistema central ejecutivo de la memoria a corto plazo (o "de trabajo", como prefirieron denominarla), al igual que el computador, podía procesar una cantidad limitada de datos de cualquier clase y controlar el sistema como un todo. El módulo viso-espacial estaba encargado de almacenar información visual o espacial a la que podía tener acceso fácilmente el sistema central (como RAM), y el bucle articulatorio se pensaba como el encargado de repasar la información verbal una y otra vez, hasta cuando el sujeto lo considerase necesario. De este modo, tal información podía usarse de inmediato o pasar a un registro más permanente (como en las cintas o discos de los computadores).

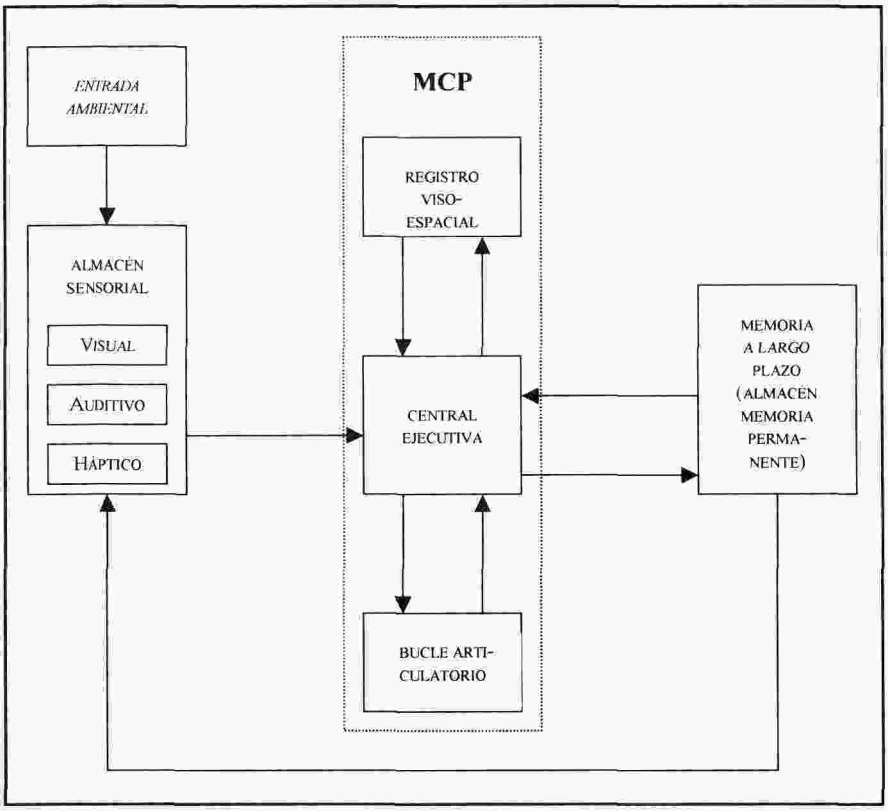

Fig. 11-2
Sistema de memoria a largo plazo y a corto plazo

Pero la teoría modular no termina con Baddeley y Hitch; aunque sus ideas motivaron la formulación de otros modelos modulares que explicaban de forma más detallada el funcionamiento de los subsistemas particulares de la memoria de trabajo, suscitaron también la creación de otras alternativas que buscaban explicar tales procesos desde perspectivas distintas. (Gathercole, 1997; Brown, 1997, y Jonides and Smith, 1997). Y prontamente, los modelos modulares se expandieron a campos distintos al de la memoria, como es el caso de los modelos de percepción visual propuestos por Marr (1982), o los de comprensión y producción de lenguaje (Manning, 1992). Incluso algunos trabajos más actuales insisten en la necesidad del modelamiento modular de ciertas funciones muy específicas, y llegan a proponer notables acercamientos entre el nivel computacional de explicación y el nivel anatómico en el que tales operaciones se instancian en el

cerebro. Un buen ejemplo de esta tentativa de investigación, son los modelos modulares propuestos a partir de las observaciones de los patrones perceptuales de la corteza olfativa y del área CA1 del hipocampo. Tales modelos no sólo explican en detalle las características operativas de los diversos subsistemas involucrados a nivel computacional, sino que buscan acercarlos a la realidad atendiendo a lo que efectivamente ocurre a nivel neuroanatómico en dichas regiones del cerebro. (Granger y Lynch, 1994. El análisis de otros modelos modulares sobre el funcionamiento de la memoria en el hipocampo puede encontrarse en: Foster, Ainsworth, Faratin and Shapiro, 1997).

11.2 Del cerebro al computador: modelos conexionistas

Cuando Marr propuso, en 1982, el marco general en el que habrían de organizarse las teorías de los modelos mentales, distinguió tres niveles distintos que deberían caracterizarse por separado en los modelos de procesamiento de información. El más alto de ellos, el "nivel computacional", especificaría los objetivos finales de la computación (percepción visual, podría ser uno de ellos). El nivel intermedio o "algorítmico" correspondería, justamente, a la especificación de los mecanismos algorítmicos que intervienen en el procesamiento que va desde la entrada (el *input*) hasta la salida (el *output*). Y, por último, estaría el nivel del *hardware*, en el que debería darse cuenta de cómo se instancia físicamente un algoritmo determinado al segundo nivel.

Sin embargo, muchos sostienen que el estudio del nivel algorítmico no puede separarse del estudio del nivel del *hardware*. Fodor, por ejemplo, sostenía que los descubrimientos alcanzados hasta el momento en el área de la neuroanatomía eran consistentes con su formulación de que los módulos se instanciaban en "arquitecturas neurales fijas". Pero los adelantos tecnológicos en imaginería cerebral y las detalladas evaluaciones de pacientes con lesiones cerebrales, han puesto en duda la posibilidad de que el estudio del procesamiento computacional de las funciones mentales esté separado de los descubrimientos al nivel físico del cerebro. Por una parte, porque los supuestos sistemas de entrada no parecen correlacionarse uno a uno con estructuras neuronales fijas, ya que otras regiones cerebrales, aparentemente desconectadas de éstas, parecen jugar algún papel en dichos procesos. Y, por otra, porque la ciencia ha

demostrado que ciertos impedimentos en los procesos que se creían específicos de un dominio particular, y por tanto independientes de los demás, alteran de diversos modos otras funciones cognitivas que, *ex hipotesis,* habrían de permanecer intactas. (Marshall, 1984).

Con el pasar del tiempo, entonces, algunos supuestos de la teoría modular de la mente se han puesto seriamente en duda, particularmente los que se refieren al encapsulamiento informativo, a la especificidad de dominio y a la idea de que todos los procesos cerebrales deban ser lineales-secuenciales. Algunos teóricos, preocupados por la aparente sin salida a la que llevan estas contradicciones entre los modelos computacionales y los descubrimientos de la neurociencia, se dieron a la tarea de construir nuevos modelos, esta vez basados en la información que poseemos sobre el cerebro, y no en el abstracto procesamiento de los computadores lineales-secuenciales tipo von Newmann. La mayor parte de estas teorías han venido a conformar otro tipo de perspectiva computacional, basada en el funcionamiento de las neuronas, y que es conocida como conexionismo o modelos de redes neuronales. Antes de entrar a examinarla, se hará un breve recuento de algunos datos de la naturaleza de las neuronas, que es bueno tener en cuenta para lograr una mejor comprensión de esta nueva alternativa computacional.

11.2.1 Las neuronas naturales

Las neuronas son las células que componen el sistema nervioso. La mayor parte de ellas son denominadas *neuronas heteropolares*, por presentar dos ramificaciones principales: el axón y las dendritas. Algunas otras, por ejemplo las células anacrinas de la retina, son *isopolares*, pues carecen de axón. No obstante, la estructura heteropolar predomina en el cerebro, aunque la forma que adoptan el axón y las dendritas depende del tipo particular al que pertenecen, de la función que cumplen y de la región cerebral en la que se encuentran. Las neuronas están organizadas en complejas redes en las que cada célula se relaciona con otra en un tipo de conexión denominada *sinapsis*, relación ésta que involucra al árbol dendrítico de una neurona A y la terminación axónica (el final del axón) de una neurona B. El árbol dendrítico y el pericarion (el punto en el que se bifurcan las dendritas) constituyen la parte receptiva de la neurona en la sinapsis. Estos elementos responden a la estimulación de otras células y

producen potenciales de membrana cuya magnitud varía de acuerdo con la intensidad de la estimulación. Dichos potenciales se atenúan a medida que se alejan del punto de estimulación, por lo cual son denominados *potenciales graduados*. En cambio, el axón responde a estimulaciones que cumplen con la "ley del todo o nada" llamadas *potenciales de acción*[1] . Cuando éstos se inician son propagados a través del axón sin sufrir atenuación alguna, hasta llegar a la terminal axónica, es decir, a la parte transmisora de la neurona. (Anadón, 1995, 17).

Aunque se conocen dos tipos de sinapsis, la química y la eléctrica, al parecer sólo la primera se presenta en el sistema nervioso de los humanos. En la sinapsis química se distinguen dos partes principales: el elemento pre-sináptico (usualmente un axón) y el postsináptico (generalmente las dendritas), los cuales se encuentran separados por una estrecha brecha de espesor homogéneo conocida como *espacio sináptico*

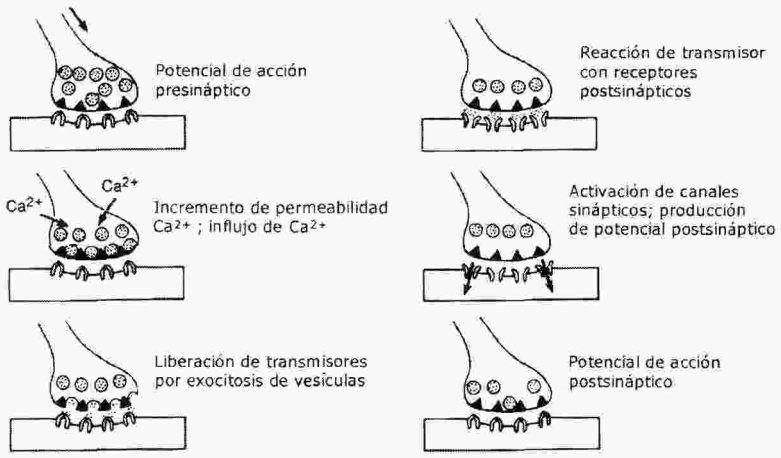

Fig. 11-3.
Esquema general de una sinapsis química (Modificado de Kandel, Siegelbaum y Schwartz, 1991, 132).

[1] Los potenciales de acción tienen un umbral de respuesta para los estímulos. Cuando un estímulo no alcanza el umbral, no se produce el impulso nervioso, pero si llega a la intensidad requerida, "el efecto en la neurona es total e irreversible y aunque se aumente la intensidad del estímulo no se producirá ya variación alguna en la magnitud del impulso. A esto se llama *ley del todo o nada*" (López, 1979, 98).

(Fig. 11-3). En las terminaciones del axón se encuentran unos pequeños botones o *vesículas sinápticas* que contienen moléculas que actúan como mediadores del impulso nervioso, que se transmite unidireccionalmente. En la membrana postsináptica aparecen, en cambio, una serie de sustancias –neuro-transmisores y neuropéptidos-, que actúan abriendo transitoriamente canales iónicos para recibir las moléculas mediadores expulsadas por la membrana presináptica y que han sido conducidas a través del espacio sináptico. El estudio de este proceso ha permitido establecer dos tipos distintos de sinapsis químicas: excitadoras e inhibidoras. En las primeras, se da una clase particular de intercambio iónico en la que el mediador despolariza la membrana postsináptica, disminuyendo su potencial de reposo (es decir, el grado de activación en el que se encuentra la neurona inactiva) y originando un potencial de acción en esta segunda neurona. En las sinapsis inhibidoras ocurre, en cambio, o bien una hiperpolarización de los iones de la membrana postsináptica (que impide el intercambio), o bien una reducción en la cantidad de las moléculas mediadoras liberadas (inhibición presináptica), lo cual impide que la segunda neurona se active. (López, 1979, 100). La naturaleza de las moléculas liberadas, de los neurotransmisores y de los neuropéptidos, varía de acuerdo con la sinapsis y el tipo de neurona, sin que aún hayan podido identificarse todas las posibles variaciones.

11.2.2 Las neuronas artificiales

A pesar de que no tuvo tanta acogida en un principio, la idea de construir modelos abstractos para explicar el funcionamiento del cerebro con base en la actividad neuronal no es nueva. Probablemente el primer intento serio en este campo fue realizado por McCulloch y Pitts (1943/1965) en su teoría de redes neuronales formales. En ésta se proponía un sistema expresado en lógica formal proposicional, cuyas reglas reflejaban ciertas propiedades de las neuronas, particularmente la "ley del todo o nada" y su capacidad de excitación e inhibición. Los descubrimientos posteriores en neurociencia y en inteligencia artificial invalidarían varios de los postulados de su teoría, por lo cual su aceptación pronto se vio disminuida. Sin embargo, a mediados de la década de los 80, una postura similar comenzó a popularizarse vertiginosamente en el ámbito de la ciencia cognitiva. Es conocida, dentro de los círculos de investigación, como

conexionismo[2] y, contrario a los modelos computacionales anteriormente vistos, basados en arquitecturas lineales-secuenciales tipo von Newmann, éste propone un acercamiento inspirado en la forma en que el cerebro mismo funciona. En palabras de uno de sus principales exponentes:

> El enfoque de PDP sugiere que debemos formular la pregunta directamente acerca del tipo de computador que el cerebro puede ser, experimentar con computadores 'tipo cerebro', y luego extraer nuestra inspiración de esos sistemas computacionales. En breve, queremos reemplazar la metáfora del computador con la metáfora del cerebro para los sistemas cognitivos. [...En general l]os modelos de PDP, como los cerebros, consisten en grandes redes de trabajo de simples unidades de procesamiento, que se comunican a través del paso de mensajes inhibitorios y excitatorios de una a otra [...] (Rumelhart, 1992, 70).

La estrategia básica de esta nueva postura es, entonces, tomar como su unidad fundamental de procesamiento algo cercano a una neurona abstracta e imaginar que las computaciones son efectuadas a través de la simple interacción entre esas unidades de procesamiento o nodos. En estos modelos, se propone que los nodos se comunican entre sí enviando números a través de las líneas que los conectan, y que ellos representan conexiones inhibitorias o excitatorias. (Rumelhart, 1989, 207). Se asume, por otra parte, que contrario a los modelos modulares, las unidades de los sistemas conexionistas no son significativas, es decir, que la información procesada en sus computaciones no es nunca una propiedad de algún nodo particular, sino que se encuentra distribuida a lo largo de todo el patrón de conexión neuronal. Además, dado que las operaciones cognitivas son sumamente rápidas a pesar de que las neuronas son bastante más lentas que los componentes de los computadores actuales, se propone que las computaciones involucran gran cantidad de procesamientos en paralelo, y no únicamente uno secuencial (de ahí que a este tipo de sistemas se les llame *procesamiento distribuido paralelo*). En breve, se puede decir que existen ocho componentes principales en los modelos conexionistas tipo PDP (Fig. 11-4):

1. Lo primero es siempre especificar un *conjunto de unidades de procesamiento,* ninguna de las cuales es una entidad significativa como tal, sino

---

[2] En lo que sigue se analizará una variable conexionista denominada *procesamiento distribuido paralelo* (PDP en adelante).

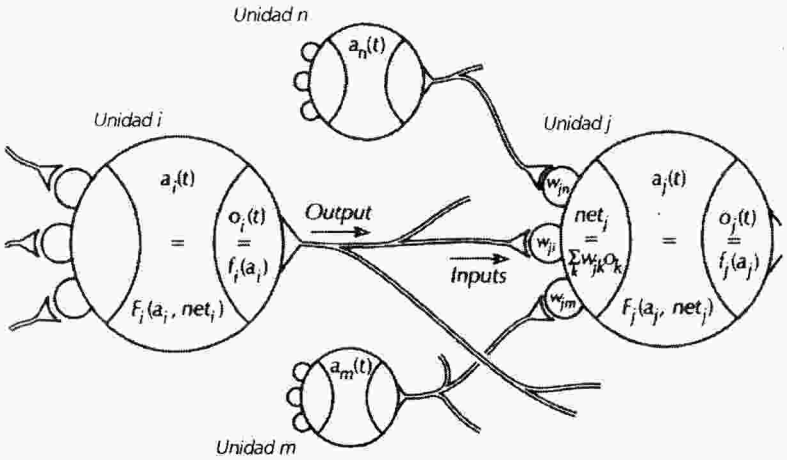

Fig. 11-4
Componentes básicos del sistema PDP. (Tomado y modificado de: Rumelhart, 1989, 210 y Rumelhart, McClelland and the PDP Research Group, 1986, 47)

que participa dentro de un gran patrón de conexión que, visto como un todo, significa algo (v.gr. una percepción visual o auditiva, o un recuerdo episódico). Lo que cada unidad hace es, simplemente, recibir una entrada (*input*) de sus unidades vecinas y, con base en una función previamente especificada para ella, computar un valor de salida que enviará a su vez a estas y/u otras unidades vecinas.

Existen tres tipos distintos de unidades (Fig. 11-5):

- Unidades de entrada: reciben datos de entrada de fuentes externas al sistema (por ejemplo, sensoriales o de sistemas externos).

- Unidades de salida: envían señales fuera del sistema (por ejemplo, al afectar sistemas motores).

- Unidades escondidas: son aquellas cuyas entradas y salidas ocurren dentro del sistema mismo, por lo cual se dice que son 'invisibles' a los sistemas externos. (Rumelhart, 1989, 211).

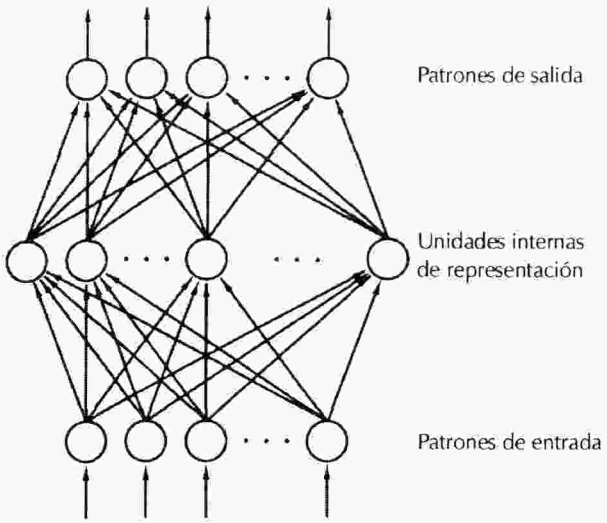

Fig. 11-5
Red de varias capas. (En: Rumelhart, 1989, 225)

Como el sistema es paralelo, se asume que puede haber tantas unidades como se quiera operando al mismo tiempo. En la figura 11-4 el conjunto de unidades (círculos) se reduce a:

$$U = \{u_i, u_j, u_n, u_m\}$$

aunque se da por sentado que puede haber muchas otras participando del sistema.

2. Hay también un *estado de activación*, representado por un vector $\mathbf{a}(t)$, que especifica el patrón de activación sobre un conjunto de unidades de procesamiento en un tiempo *t*. (Rumelhart, McClelland and the PDP Research Group, 1986, 48)[3]. De este modo, es el patrón de activación total de todo el sistema de unidades lo que captura el significado de aquello que éste

[3] Es útil valerse de la notación del álgebra vectorial en los modelos PDP. En términos generales, un vector (representado por una letra minúscula en negrita) es una manera económica de describir patrones de números. Así, por ejemplo, el vector correspondiente al patrón numérico cuyos componentes son la edad, la altura y el peso de María, puede representarse (Rumelhart, McClelland and the PDP Research Group, 1986, capítulo 9):

$$\text{María} \begin{vmatrix} 10 \\ 132 \\ 30 \end{vmatrix}$$

representa en determinado momento. Los valores de activación pueden ser continuos o discretos, y es usual restringirlos al conjunto de números reales comprendidos entre [0,1], de forma que 1 indicará la total activación de un nodo, y 0 su total inactividad. (En algunos casos, se toman valores entre [-1,1]).

3. Debe especificarse también una *función de salida* para cada unidad. Dado que éstas interactúan con otras transmitiéndoles señales, la fuerza de esas señales –y por tanto, el grado en que afectan a sus vecinas- está determinada por su nivel de activación. (Rumelhart, 1989, 212). De ahí que a cada unidad u_j le corresponda una función $f_i(a_i(t))$ que convierte el estado de activación en una función de salida:

$$o_i(t)$$

Existen tres tipos principales de funciones de salida: 1) de identidad (el resultado de la función es idéntico al estado de activación), 2) de umbral (una unidad A sólo tiene efecto sobre una B si su *input* hace que su estado de activación exceda un cierto umbral numérico), y 3) estocástica (el *output* depende de ciertos resultados probables del estado de activación).

4. La excitación o inhibición de una conexión entre dos unidades está representada por el *peso de conexión*, el cual usualmente se indica con un número real que varía entre –1 y 1. Así, por ejemplo,

$$w_{ji}$$

representa el peso de conexión de la unidad u_i a la unidad u_j. Si es un número positivo significa que u_i excita a u_j, si es negativo u_i inhibirá a u_j, y si es 0 será porque u_i no está en directa relación con u_j.

5. El quinto elemento es el *patrón de conectividad*. Este constituye propiamente lo que el sistema 'sabe' -es decir, el tipo de información que codifica-, y determina cómo se comportará ante una entrada arbitraria. En la mayor parte de modelos, la entrada (*input*) total de una unidad es la suma ponderada de las entradas separadas de cada una de las unidades que se conectan a ella. (Rumelhart, 1989, 212). Esto es: los *inputs* de todas las unidades de entrada son multiplicados por sus respectivos pesos de conexión, y luego estos resultados se suman para obtener la entrada total de una unidad. Es usual representar estos patrones de conexión con matrices de peso (**W**) que relacionan las unidades de entrada, las unidades de salida y sus respectivos pesos de conexión.

6. Debe especificarse una *regla de propagación* que tome los datos de salida en un determinado momento o(t) y los combine con matrices de conexión para producir una entrada-red (*net input*) para cada tipo de *input* en una unidad. (Rumelhart, McClelland and the PDP Research Group, 1986, 51). Así, por ejemplo, cuando los *inputs* son excitatorios, la regla de propagación será el producto vectorial:

$$net_e = W_e\ o(t)$$

7. Se necesita también establecer una *regla de activación* que indique la manera en que diferentes entradas a una unidad u_i se combinan, tanto entre sí, como con el actual estado de activación de u_i, para producir un nuevo estado de activación. (Rumelhart, McClelland and the PDP Research Group, 1986, 51). Esta regla se representa como una función F_i sobre a_i (t) y las entradas-red, para generar un nuevo estado de activación:

$$F_i = net_i = S_j\ w_{ij}\ o_j\ (t)$$

Como se anotó antes, F_i puede ser una función de identidad, de umbral o estocástica.

8. Finalmente, es necesario recaer en otra propiedad de las redes de PDP: éstas son capaces de aprender -es decir, de cambiar su estructura de *conocimiento-, modificando los patrones de interconexión entre sus nodos*. En general, estas modificaciones son de tres tipos (Rumelhart, 1989, 213):

- desarrollo de nuevas conexiones,
- pérdida de conexiones existentes o
- modificación de fuerzas de conexiones existentes.

El aprendizaje de la red, producto de alguna de estas modificaciones, puede ser determinado por una *regla de aprendizaje*, una variable de la ley neurológica de Hebb: "si una unidad u_i recibe una entrada de otra unidad u_j en un tiempo en que ambas unidades están altamente activas, entonces el peso w_{ij} a u_i de u_j debe ser fortalecido". (Rumelhart, 1989, 214). Lo cual puede representarse, atendiendo a las variables temporales, por medio de la fórmula:

$$\delta w_{ij} = g(a_i\ (t), \tau_i\ (t))\ .\ h(o_j\ (t), w_{ij})$$

donde δ indica la proporción en el cambio de la conexión de u_j a u_i dada por el producto de la función g, entre el estado de activación de la unidad en un

tiempo dado ($a_i(t)$) y una entrada de enseñanza (τ_i), y una función *h*, entre el valor de salida de u_j (o_j) y el peso de conexión entre u_j y u_i (w_{ij}).

Una de las muchas variables de la regla de aprendizaje es el *algoritmo de retro-propagación* (*back-propagation algorithm*), un medio para producir un valor de salida conforme a un estado deseado. Consiste, en pocas palabras, en que, para cualquier entrada, el sistema es informado acerca de qué *output* o salida debe darse. Esta información es comparada con el estímulo actual de la red –cuando el estímulo se presenta-, y se realiza el ajuste. De este modo, aunque se necesita exponer al sistema a varios estímulos, llega un momento en que aprende cuál es el *output* correcto. (Parkin, 1996). Precisamente, el hecho de que los modelos de tipo PDP se consideren redes de aprendizaje activo -inspiradas en el procesamiento neuronal-, ha llevado a que muchos neuropsicólogos se interesen por estudiar las consecuencias de los daños en los modelos, para extraer así conclusiones relativas al funcionamiento del cerebro humano. Por ello, se han construido muchas instancias de redes PDP que, al ser afectadas, producen patrones de respuesta muy similares a los efectuados por pacientes con el déficit correspondiente.

Poco a poco los modelos conexionistas han ganado terreno en el ámbito de la investigación en psicología cognitiva, por un lado, porque resultan ser muy buenos candidatos para resolver viejos problemas cognitivos de imposible resolución dentro de los modelos de procesamiento lineal (como la *búsqueda del mejor pareamiento* o la *degradación elegante*. Ver: Rumelhart, 1989, 215ss; y Rumelhart, McClelland and the PDP Research Group, capítulo 4). Y por otro, porque su cercanía con la realidad neuronal los ha postulado como una excelente estrategia para modelar un amplio espectro de procesos cognitivos, que se extiende desde la mera percepción visual hasta la comprensión del lenguaje y la resolución de problemas. (Smolensky, 1989, 233). Estudiar cada uno de estos modelos excede los límites de este capítulo, por lo cual se analizará de forma muy somera un modelo de la memoria, propuesto por McClelland y Rumelhart (1986, capítulo 17), con el objeto de ejemplificar los elementos de este nuevo enfoque computacional.

11.2.3. Memoria y conexionismo

Uno de los problemas centrales de los modelos de memoria tiene que ver con la manera en que aprendemos información general a partir de conocimientos específicos. La mayor parte de los modelos de aprendizaje-memoria de tipo linear-secuencial sostenían que nuestro conocimiento general era producto de una suerte de abstracción de percepciones particulares de eventos ejemplares. Así, aprendíamos a reconocer instancias de la palabra "perro" luego de haberla visto *n* veces y de haber abstraído, de algún modo, sus características esenciales. Y se decía, de modo similar, que conocíamos la palabra "perro" porque nuestra experiencia había hecho que ésta se codificase en una estructura neuronal fija en la MLP. Sin embargo, quienes sostenían esta clase de modelos, se veían en dificultades a la hora de dar cuenta de la naturaleza de dicha abstracción, así como al momento de responder por lo que ocurriría si esa estructura neuronal fija que codificaba "perro" llegara a fallar.

Pues bien, el PDP propone otra alternativa: en lugar de suponer que la información se almacena en lugares determinados (como la MLP o la MCP), se asume que se encuentra distribuida en patrones de conexión de unidades homogéneas, del tipo que se explicó en el punto anterior. Tales unidades se encuentran organizadas en módulos (Fig. 11-6) que reciben y envían información a otros módulos, y en cada uno de los cuales puede haber miles de millones de unidades. (McClelland, Rumelhart and the PDP Research Group, 1986, 174).

Esto implica, por un lado, que cada módulo combina diferentes fuentes de información (i.e. no hay especificidad de dominio) y, por otro, que cada unidad puede estar involucrada en la representación de diferentes cosas. "Así, por ejemplo, usted tiene una amiga llamada Rebecca, quien es una estudiante de postgrado, de 26 años, casada y sin hijos. De acuerdo con el modelo [...], cada una de esas piezas de información acerca de Rebecca es almacenada en un patrón [*sic*] separado" (Eysenck y Keane, 1995, 152). En consecuencia, al hablar de representaciones distribuidas, los modelos de PDP pueden evitar caer en la difícil explicación de la abstracción de los modelos lineares-secuenciales, arguyendo simplemente que las distintas presentaciones de, verbigracia, la palabra "perro", conducen a la activación de patrones de diferentes perros conocidos, los cuales comparten características prototípicas de estos animales. (Eysenck y Keane, 1995, 153).

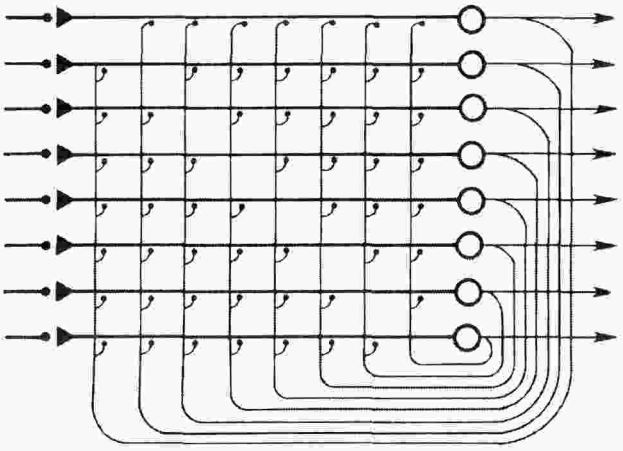

Fig. 11-6
Módulo simple de procesamiento de información con ocho nodos. Cada nodo recibe información de otros módulos y envía señales de salida a otros más. Las ramas internas indican conexiones intra-módulo. (En: McClelland, Rumelhart and the PDP Research Group, 1986, 174)

Que la información esté distribuida en diversos patrones neuronales, representa otra ventaja sobre los modelos lineares, ya que implica que el sistema podría funcionar relativamente bien aún cuando una unidad -o un grupo de unidades- se encontrara alterada. Así, por ejemplo, para recuperar información relativa a Rebecca, se pueden emplear varios caminos simultáneamente: "amigos casados", "amigos estudiantes de postgrado", "conocidos sin hijos", etc. De esta forma, si por algún motivo la vía "amigos estudiantes de postgrado" se encuentra afectada, de acuerdo con el modelo, se podría llegar al mismo resultado (Rebecca) a través de otros caminos alternos. En un modelo linear-secuencial, en cambio, la interrupción en una vía significa la detención de toda la computación.

Sin embargo, que el conocimiento en memoria se encuentre distribuido, no significa que ese patrón haya de permanecer activo siempre, una vez se ha adquirido tal información. En realidad, los patrones de conexión "van y vienen", dejando *rastros* cuando han pasado. (McClelland, Rumelhart and the PDP Research Group, 1986, 176). Estos rastros deben entenderse como cambios en los pesos de conexión entre las unidades, y el rastro de un patrón particular de memoria será el conjunto total de cambios en los pesos de conexión de un módulo. Este conjunto total de cambios se denomina *incremento* y su valor se obtiene con la sumatoria

de los resultados de los cambios de cada peso particular de conexión, los cuales, a su vez, están determinados por una variable de la regla de Hebb llamada *regla delta* (análisis detallados de esta regla pueden encontrarse en: Rumelhart, McClelland and the PDP Research Group, 1986, capítulos 2, 8 y 11). Ello, por supuesto, hace que múltiples exposiciones a determinada información fortalezca rastros particulares o que la falta de estímulos los debilite.

Todo esto permite lo que se ha llamado *generalización espontánea*, un concepto que involucra el "recuerdo" de información que no se ha aprendido explícitamente. Es decir, puesto que el proceso de recobro no tiene una dirección fija, los conocimientos previamente adquiridos pueden participar en la deducción de otros que no han sido aprendidos por experiencia directa. Por ejemplo, alguien puede preguntarle a usted si los conservadores son en promedio más viejos que los liberales. Aunque nunca haya estado en presencia de un *input* que afirme o niegue explícitamente tal proposición, usted podría, basándose en su conocimiento de conservadores ancianos, generalizar una respuesta y acertar. De modo similar, el modelo permite rellenar información faltante al respecto de un individuo, objeto o evento sobre la base de información conocida de individuos, objetos o eventos similares. Suponga que olvida que Rebecca tiene 26 años. Cuando, en su intento por recordarlo, activa ciertas otras características de Rebecca, se activan también otros patrones de información sobre individuos similares a ella. Si estos se encuentran entre los 20 y los 30 años, la respuesta correcta podría aparecer o, al menos, una buena aproximación. Este proceso se denomina *asignación por omisión.* (Eysenck y Keane, 1995, 152).

La evidencia empírica ha fortalecido la fe en este modelo y ha motivado, también, la creación de sistemas conexionistas que buscan explicar la naturaleza de ciertas amnesias. (McClelland, Rumelhart and the PDP Research Group, 1986, capítulo 25). No obstante, hay quienes se resisten a creer que la cercanía con el cerebro sea suficiente garantía para explicar los procesos mnésicos y arguyen, en cambio, que para efectos prácticos los modelos modulares siguen estando a la vanguardia (algunos ejemplos de este debate pueden encontrarse en: McNaughton and Smolensky, 1991; Shapiro and Eichenbaum, 1997; Shanks, 1997). Las opiniones al respecto del futuro del modelo conexionista en neuropsicología cognoscitiva, están bastante divididas. Por un lado, hay un grupo grande de investigadores que consideran que esta postura es bastante promisoria, y dedican sus estudios a la formu-

lación de diversos modelos que, basándose en ella, explican la naturaleza de muchos desórdenes cognitivos adquiridos. Otros, más escépticos, consideran que el conexionismo tiene muchas limitaciones, tanto teóricas como prácticas, que lo debilitan como la mejor propuesta de modelamiento computacional. (Reggia, Berndt and D'Autrechy, 1994). ¿Cuál podría ser entonces la alternativa?

11.3 La verdad relativa de los modelos computacionales

Se podría decir, según se ha visto, que la neuropsicología cognoscitiva es una rama de investigación de la neuropsicología, que se sirve tanto de los resultados clínicos de las experiencias con pacientes que han sufrido daño cerebral, como de los adelantos teóricos y técnicos de la ciencia cognitiva (ver especialmente en: Churchland y Sejnowski, 1991, 1992). Esto, por supuesto, le trae grandes ventajas y le permite proponer explicaciones muy completas sobre el funcionamiento normal del cerebro humano. No obstante, la convierte también en heredera de importantes problemas prácticos y teóricos.

A nivel práctico, por ejemplo, la neuropsicología cognoscitiva debe enfrentar la crítica de quienes consideran el estudio de caso único –indispensable para la formulación de modelos computacionales-, como una manera poco científica de acercarse a la realidad de las funciones psicológicas. (Shallice, 1991; Parkin, 1996). En su lugar, le proponen adoptar un tipo de investigación grupal, sobre la base de que los procesos cognitivos estudiados en distintos sujetos son homogéneos. Este supuesto claramente es difícil de aceptar, pues los recientes avances en neuroimaginería sugieren que tanto a nivel anatómico, como a nivel de procesamiento neuronal, incluso habilidades tan básicas como la percepción o la memoria pueden ser efectuadas de maneras cualitativamente distintas en varios individuos. (Parkin, 1996).

Así mismo, a nivel teórico surgen para ella otras dificultades que tendrá que sortear para llegar a ser considerada una buena estrategia para la explicación de la cognición. En primer lugar, deberá atender al hecho de que los fenómenos cognoscitivos son de una muy peculiar naturaleza, pues, por un lado, son usualmente inaccesibles a la observación directa o a la introspección y, por el otro, abarcan un rango que va desde simples escogencias (como presionar o no un botón), "hasta complejos

grupos de actividades de resolución de problemas tales como controlar incendios, volar aeroplanos o navegar en mar abierto". (Seifert y Shafto, 1994, 410). En este sentido, se dice que la neuropsicología cognoscitiva habrá de tener la suficiente *cobertura* como para dar cuenta de los más diversos dominios de la cognición humana. Además, en tanto que se vale metodológicamente de los modelos computacionales para formular sus explicaciones, habrá de procurar que estos puedan aplicarse a cualquier clase de alteración neuropsicológica.

Por otra parte, sus modelos habrán de ser *adecuados, precisos y susceptibles de revisiones y refinamientos*, en la medida en que la experiencia así lo requiera. Pero esto sólo es posible si dichos modelos dependen de "unos pocos procesos fundamentales, estructuras y reglas de composición; si estos procesos y estructuras son fácilmente, aunque quizás indirectamente, correlacionados con fenómenos concretos, observables; y si los métodos para formular, comprobar y refinar modelos son directos". (Seifert y Shafto, 1994, 414). Así, la credibilidad de la teoría empezará a garantizarse y su aceptación aumentará. No obstante, algunos teóricos –como lo señala Parkin, siguiendo a Roediger- están preocupados por la ausencia de un principio organizativo más rígido en las teorías computacionales. Afirman, entre otras cosas, que se han descubierto tantas disociaciones modulares tan sumamente específicas, que las teorías computacionales de la cognición podrían correr la misma suerte que la psicología de la Gestalt, en la que, en algún momento, llegaron a tenerse hasta 114 leyes distintas de la percepción. (Parkin, 1996).

Finalmente, es necesario que la teoría sea *realista*, esto es, que mantenga consistencia con el resto del conocimiento científico. En este sentido, una buena teoría cognoscitiva deberá "salvar la brecha" entre el nivel algorítmico de análisis y el nivel cerebral –o de *hardware*-, a través de una suerte de mapeo entre los dos. (Seifert y Shafto, 1994, 413). Como se vio, esta crítica ha sido esgrimida principalmente en contra de las teorías de procesamiento simbólico y modular, pues arguyen que sus postulados contradicen observaciones de la neurociencia. (Marshall, 1984). Sin embargo, hay quienes también critican al conexionismo por su falta de consistencia a la hora de clasificar las diversas alteraciones neuropsicológicas que simulan, ya que las 'etiquetas' que utilizan para denominarlas (v.gr. *afasia agramática* o *dislexia profunda*) "frecuentemente presentan una imagen errónea de la homogeneidad sintomática entre pacientes". (Reggia, Berndt and D'Autrechy, 1994, 327).

Inevitablemente, surge aquí la siguiente pregunta: ¿cuál puede ser la solución para estos problemas? Que, dado el contexto, sólo podrá ser contestada si se encuentra una respuesta para esta otra: ¿cómo debe ser, en últimas, una buena teoría de los modelos mentales? Muchos sostienen que el objetivo de una teoría computacional de la mente debe ser proveernos de un solo modelo arquitectónico-funcional de la cognición, lo suficientemente amplio, coherente y realista como para explicar todas las tareas cognitivas. Esta opción, en todo caso, se ha puesto seriamente en duda y cada vez más aparecen teóricos que proponen, en cambio, una teoría mixta, en la que se utilice el procesamiento lineal-secuencial para modelar ciertos procesos, y el conexionismo para modelar ciertos otros (esta posibilidad se examina en Hadley, 1999). Pero esto, nuevamente, nos remite al problema de la ausencia de un principio regulativo y la historia de la ciencia nos ha demostrado los problemas a los que se enfrentan las teorías que carecen de un *corpus* que unifique sus ideas. De hecho, es también probable que ninguno de los modelos hasta ahora diseñados se convierta en la solución al problema. Tal vez deban idearse otros que comprendan ciertos aspectos de la cognición humana que los actuales olvidan, como el carácter social, cultural y evolutivo del hombre. En cualquier caso, lo único que realmente puede afirmarse con total certeza es que aún hay mucho trabajo por hacer en estas áreas.

Bibliografía

Anadón, R. (1995). Estructura de las neuronas naturales. En: Barro y Mira, 1995.

Atkinson, R. C. y Shiffrin, R. M. (1968). Human Memory: A proposed system and its control processes. En: Spence, K.W. y Spence J.T. (eds.) "The psychology of learning and motivation". Vol. 2. Academic Press, London.

Baddeley, A. D. (1995). The Psychology of Memory. En: "Handbook of Memory Disorders". A.D. Baddeley, B.A. Wilson and F.N. Watts (eds.).

Barro, S. y Mira, J. (eds.) (1995). Computación neuronal. Universidade de Santiago de Compostela.

Boden, M. A. (1977). Artificial Intelligence and Natural Man. Harvester, Sussex.

Brown, G. D. A. (1997). Formal models of memory for serial order: A review. En: Conway, 1997.

CHURCHLAND, P. S. AND SEJNOWSKI, T. J. (1991). Perspectives on Cognitive Neuroscience. En: Lister and Weingartner, 1991.

CHURCHLAND, P. S. AND SEJNOWSKI, T. J. (1992). The Computational Brain. MIT Press, Cambridge, Mass.

CONWAY, M. A. (ed.) (1997). Cognitive Models of Memory. MIT Press, Cambridge, Mass.

DAVIS, S. (ed.) (1992). Connectionism: Theory and Practice. Oxford University Press, NY.

EYSENCK, M. W. AND KEANE, M. T. (1995). Cognitive Psychology. A Student's Handbook. 3rd Edition. Psychology Press, UK.

FODOR, J.A. (1986). La modularidad de la mente. Morata, Madrid. (Título original: "The Modularity of Mind". MIT Press, 1983)

FOSTER, J.; AINSWORTH, J.; FARATIN, P. AND SHAPIRO, J.. (1997). Implementing a mathematical model of hippocampal memory function. En: Conway, 1997.

GATHERCOLE, S. E. (1997). Models of verbal short-term memory. En: Conway, 1997.

GRANGER, R. Y LYNCH, G. (1994). Cognitive modularity: computational division of labor in the brain. En: Hendler, 1994.

HADLEY, R. F. (1999). Connectionism and Novel Combinations of Skills: Implications for Cognitive Architecture. En: "Minds and Machines", 9: 197-221.

HAUGELAND, J. (ed.) (1997). Mind Design II. MIT Press, Cambridge, Mass.

HENDLER, J.A. (ed.) (1994). Computational Modelling and Neuropsychology. En: Handbook of Neuropsychology. Vol. 9. F. Boller and J. Grafman (eds.).

JONIDES, J. AND SMITH, E. E. (1997). The architecture of working memory. En: Rugg, 1997.

JOHNSON-LAIRD, P. N. (1998). The Computer and the Mind. Harvard University Press, Cambridge, Mass.

KANDEL, E. R.; SCHWARTZ, J. H.; JESSELL, T. M. (1991). Principles of neural science. Appleton & Lange, Norwalk, CT.

KANDEL, E. R.; SIEGELBAUM, S. A.; SCHWARTZ, J. H. (1991). Synaptic Transmission. En: Kandel, Schwartz y Jessell, 1991.

LISTER, R. G. AND WEINGARTNER, H. J. (eds.) (1991). Perspectives on Cognitive Neuroscience. Oxford University Press, NY.

LÓPEZ, L. (1979). Anatomía funcional del sistema nervioso. México: Limusa.

MANNING, L. (1992). Introducción a la neuropsicología clásica y cognitiva del lenguaje. Ed. Trotta, Valladolid.

MARR, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society*, B207: 187-217. Londres.

MARR, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: W. H. Freeman.

MARSHALL, J. C. (1984). Multiple perspectives on modularity. En: "Cognition", 17, No. 1: 209-242.

MCCULLOCH, W. S. Y PITTS, W. H. (1943/1965). A Logical Calculus of the Ideas Inmanent in Nervous Activity. En: W.S. McCulloch, *Embodiments of Mind*. Cambridge, MA: MIT Press, 19 – 39.

MCCLELLAND, J. L., RUMELHART, D. E., AND THE PDP RESEARCH GROUP. (1986). Parallel Distributed Processing. Explorations in the Microestructure of Cognition. Vol. 2: Psychological and Biological Models. MIT Press, Cambridge, Mass.

MCNAUGHTON, B. AND SMOLENSKY, P. (1991). Connectionist and Neural Modeling: Converging in the Hippocampus. En: Lister and Weingartner, 1991.

NEWELL, A. Y SIMON, H. A. (1972). Human Problem Solving. Englewood Cliffs, N.J.: Prentice Hall.

NEWELL, A. Y SIMON, H. A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the Association for Computing Machinery,* 19. (Reimpreso en: Haugeland, 1997).

PARKIN, A. (1996). Explorations in Cognitive Neuropsychology. Blackwell, Oxford.

POSNER, M. I (1978). Chronometric explorations of mind. Erlbaum.

REGGIA, J. A.; BERNDT, R. S. AND D'AUTRECHY, L. (1994). Connectionist models in neuropsychology. En: Hendler, 1994.

RUGG, M. D. (ed.) (1997). Cognitive Neuroscience. Psychology Press, UK.

RUMELHART, D.E., MCCLELLAND, J. L., AND THE PDP RESEARCH GROUP. (1986). Parallel Distributed Processing. Explorations in the Microestructure of Cognition. Vol. 1: Foundations. MIT Press, Cambridge, Mass.

RUMELHART, D. E. (1989). The Architecture of Mind: A Connectionist Approach. En: Haugeland, 1997.

RUMELHART, D. E. (1992). Towards a Microestructural Account of Human Reasoning. En: Davis, 1992.

TURING, A. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society,* 42: 230-265.

SHALLICE, T. (1991). From neuropsychology to mental structure. En: "Behavioral and Brain Sciences", 14: 429-469.

SHAPIRO, M. L. AND EICHENBAUM, H. (1997). Learning and memory: computational principles and neural mechanisms. En: Rugg, 1997.

SHANKS, D. R. (1997). Representation of categories and concepts in memory. En: Conway, 1997.

SEIFERT, C. M. AND SHAFTO, M. G. (1994). Computational models of cognition. En: Hendler, 1994.

SMOLENSKY, P. (1989). *Connectionist Modeling: Neural Computation/Mental* Connections. En: Haugeland, 1997.