

## LA ESTRUCTURA DE LAS REPRESENTACIONES MENTALES: UNA PERSPECTIVA INTEGRADORA\*

MANUEL LIZ

### 1. La realidad psicológica de las representaciones mentales y el problema de su estructura característica

No siempre se ha admitido que la noción de representación mental sea suficientemente coherente como para ser incorporada en nuestras teorías sobre la mente. En esta línea, tanto el conductismo lógico como el conductismo psicológico abandonan gustosamente dicha noción. Lo mismo cabe decir de otros eliminativismos de tipo neurofisiológico, computacional, etc.

Y, aun preservando para la noción de representación mental algún papel teórico, tampoco todos admitirían que las representaciones mentales sean entidades psicológicamente reales. A pesar de la apariencia paradójica de esta afirmación, las representaciones mentales podrían resultar no ser entidades claramente detectables y localizables en los sujetos. El externalismo en la determinación del contenido de las representaciones mentales sugiere esto. Hasta tal punto que incluso, según algunos autores, se pondría en peligro nuestro propio autoconocimiento. Apelar a un mundo de representaciones mentales podría también tener, por ejemplo, un valor meramente heurístico. Asimismo, las representaciones mentales podrían ser propiedades sólo atribuidas a efectos de conseguir, en ciertas condiciones, una mayor capacidad predictiva y de control sobre el comportamiento de determinados sistemas. O, contra todo pronóstico griceano, las actitudes proposicionales podrían resultar ser analizables como cierto tipo de relaciones que un sujeto psicológico mantiene con ciertos objetos portadores de propiedades semánticas que son siempre externos a él mismo y que, en definitiva, son ciertas expresiones de un

\* Una versión previa de este trabajo fue presentada en el Coloquio SADAF/1994. Quisiera destacar aquí el estimulante ambiente filosófico que uno puede encontrar en aquellas tierras, lejanas sólo en los mapas, y agradecer a todos los participantes de dicho Coloquio sus comentarios y certeras observaciones. Espero que, entre otras, algunas de las cuestiones suscitadas por Fernando Broncano, Antoni Gomila, Marcelo Sabatés, Hugo Saravia y David Sosa encuentren ahora una respuesta más matizada.

lenguaje natural.

El caso es que, aun admitiendo la relevancia teórica de las representaciones mentales y su realidad psicológica, aun admitiendo que exista un especial nivel de análisis desde el cual deba aceptarse la realidad de las representaciones mentales en los sujetos, también podría no admitirse que todas las representaciones mentales deban tener esencialmente una estructura peculiar y específica. Se puede ser representacionalista sin postular ninguna estructura peculiar y específica que deba ser exigida de manera esencial a todas las representaciones mentales en general. Tal vez, simplemente, las representaciones mentales en general no tengan por qué manifestar necesariamente ninguna estructura característica, ninguna estructura esencial, compartida por todas las representaciones mentales, sin la cual no sean posibles estados y procesos mentales como los nuestros. Tal vez haya siempre, en principio, una variedad indefinida de estructuras capaces de producir los mismos resultados. Pero, lo que es más importante, puestos a hacer intervenir otras consideraciones más empíricas a fin de seleccionar, entre todos esos candidatos potenciales, una única estructura para ser postulada como psicológicamente real, tal vez tengamos que reconocer que no todas las representaciones mentales exigen lo mismo, que sólo algunas representaciones mentales requieren algunos tipos peculiares de estructuras y que, además, puede que algunas de esas estructuras sólo sean posibles si también se tienen otras.

No obstante, dos de las posiciones más influyentes hoy día en el campo de las ciencias cognitivas, el computacionalismo clásico y el conexionismo no eliminativista, sí aceptan

(1) que las representaciones mentales son entidades psicológicamente reales en los sujetos, y

(2) que las representaciones mentales son objetos con una estructura característica.

Y las discusiones filosóficas entre ambas posiciones se centran en el problema de cuál deba ser esa estructura. Algunos fragmentos de Fodor y Pylyshyn (1988, pp. 11-12) pueden servir como muestra de esto:

In short, the issue between Classical and Connectionist architecture is not about the explicitness of rules; as we'll presently see, Classical architecture is not, per se, committed to the idea that explicit rules mediate the etiology of behavior. And it is not about the reality of representational states; Classicists and Connectionists are all Representational Realists. And it is not about

nonrepresentational architecture...

So, then, what is the disagreement between Classical and Connectionist architecture about?

[...] Classicists and Connectionists all assign semantical content to something. Roughly, Connectionist assign semantical content to "nodes" (that is, to units or aggregates of units) —i.e., to the sorts of things that are typically labeled in Connectionist diagrams; whereas Classicists assign semantic content to expressions —i.e., to the sorts of things that get written on the tape of Turing machines and stored at addresses in von Neumann machines. But Classical theories disagree with Connectionist theories about what primitive relations hold among these content-bearing entities. Connectionist theories acknowledge only causal connectedness as a primitive relation among nodes [...]. By contrast, Classical theories acknowledge not only causal relations among the semantically evaluable objects that they posit, but also a range of structural relations, of which constituency is paradigmatic.

Tanto el computacionalismo clásico como el conexionismo no eliminativista se aferran a la tesis de que las representaciones mentales en general deben tener esencialmente una estructura característica. Y reducen el problema a discutir si esa estructura es, en el sentido más excluyente posible, clásica o conexionista. Sin embargo, tal vez sea justamente esa tesis de la estructura característica la que haya que abandonar.

En este trabajo, intentaré defender que las representaciones mentales en general no tienen de manera esencial ninguna estructura característica. O, al menos, que no hay razones concluyentes para suponer que la tienen en general y de forma incondicionada. En particular, defenderé que la estructura simbólica basada en relaciones sintácticas de algunas representaciones mentales sólo surge bajo una serie de condiciones muy especiales entre las que hay que incluir el que tengan también otras estructuras no simbólicas.

Esto nos obligará a considerar los sistemas conexionistas como algo más que simples implementaciones de los sistemas clásicos y a defender una perspectiva integradora.

Examinaremos algunos de los argumentos principales del computacionalismo clásico y del conexionismo no eliminativista y, a partir de ellos, defendemos nuestra posición. Pero antes, digamos algo más sobre el representacionalismo.

## 2. Representacionalismo mínimo y algunas distinciones posteriores

En un sentido muy mínimo, el representacionalismo afirmarí­a que para

tener una teoría adecuada de la mente es necesario postular en los sujetos estados mentales representacionales.

Ciertos estados de la mente, identificables de alguna manera con ciertas clases de estados físicos (presumiblemente del cerebro en nuestro caso biológico), representan estados del mundo y son semánticamente caracterizables en términos de esos estados del mundo que representan. Lo que una representación representa es su contenido semántico. Las representaciones mentales quedan caracterizadas a través de sus contenidos semánticos. Y se describe el contenido semántico de una representación mental describiendo el estado del mundo que representa.

Las representaciones mentales son entidades intencionales. Son las piezas básicas de la intencionalidad. Al representar algo, una representación mental se dirige muy selectivamente hacia un objeto peculiar que puede tanto existir realmente como no existir. La existencia ni quita ni pone nada a estos objetos. Los objetos hacia los que se dirigen las representaciones mentales son sus contenidos semánticos. El contenido semántico de una representación mental es su objeto intencional. Lo representado por una representación mental, su contenido semántico y su objeto intencional coinciden.

El carácter tremendamente selectivo de las representaciones mentales y su indiferencia respecto de la existencia real de lo representado hacen que las descripciones de las representaciones mentales se comporten de una manera muy especial. Dichas descripciones generan contextos intensionales en los que falla la sustitución de equivalentes lógicos y la generalización existencial.

Lo anterior formaría parte del representacionalismo mínimo. Ciertamente, declararse representacionalista en este sentido mínimo no es decir gran cosa. Sin embargo, las representaciones mentales podrían ser de muchos tipos. Y aquí ya hay varias distinciones importantes que cabría hacer. Son las siguientes:

(1) \* Representaciones mentales desarrolladas a un nivel conceptual.

\* Representaciones mentales desarrolladas a un nivel subconceptual.

(2) \* Representaciones mentales con relevancia cognitiva.

\* Representaciones mentales sin relevancia cognitiva.

(3) \* Representaciones mentales con una estructura característica simbólica del tipo general propuesto por el computacionalismo clásico.

\* Representaciones mentales con una estructura característica no simbólica del

tipo general propuesto por el conexionismo no eliminativista.

- \* Representaciones mentales con otras diferentes estructuras características.
- \* Representaciones mentales sin ninguna estructura característica en especial.

A continuación, explicaré estas distinciones. Al hacerlo, no haré más que recoger ciertos usos ya establecidos. Sin embargo, el tenerlas presentes de un modo claro y preciso permitirá un mejor planteamiento de los problemas.

### 3. Nivel conceptual y nivel subconceptual

Las representaciones mentales pueden desarrollarse a un nivel conceptual o a un nivel subconceptual. Digamos que se desarrollan en el primer nivel si su descripción semántica en términos de lo que representan hace uso de conceptos no fácilmente reducibles a otros conceptos expresables en un vocabulario físico (en alguna extensión razonable del lenguaje de la física). Y que se desarrollan en el segundo nivel si su descripción semántica sí hace uso de conceptos expresables en un vocabulario físico (nuevamente, en alguna extensión razonable del lenguaje de la física) o resulta plausible una reducción tal.

Una representación del avión en el que he viajado a Argentina aterrizando suavemente en el aeropuerto de Buenos Aires es una representación mental de nivel conceptual. Una representación de Alfonsín dando la mano a Felipe González también. Así como una representación de la palabra "mesa" o de cualquier otro símbolo. Pero podemos fácilmente asumir que una representación de que la temperatura de la habitación esté ahora mismo subiendo no sea una representación mental desarrollada a un nivel conceptual. Ni, tampoco, una representación de una sustancia de sabor amargo. Y tampoco lo sería, en general, la representación de cualquier objeto físico como teniendo una determinada forma, dimensiones, color, peso, sabor, etc.

Las que estamos llamando representaciones mentales desarrolladas a un nivel conceptual no son simplemente identificables con las representaciones mentales atribuidas desde una "perspectiva intencional", desde un nivel "ecológico" o desde un nivel de análisis "semántico". Nuestra distinción entre un nivel conceptual y uno subconceptual se traza en relación con el tipo peculiar de contenido semántico presente en las representaciones y con la posibilidad de describirlo adecuadamente en un lenguaje cercano al lenguaje físico. Es, por lo tanto, una distinción interna a lo que podemos encontrar desde esos otros puntos de vista.

Las representaciones mentales de nivel subconceptual, por otro lado, tienen las mismas propiedades intencionales que el resto de las representaciones mentales. En particular, sus descripciones deben seguir originando contextos intensionales donde quede puesta en cuestión la sustitución de equivalentes lógicos y la generalización existencial. Su descripción en términos cercanos a lo físico no puede hacernos olvidar esto.

Podríamos ser representacionistas admitiendo sólo la existencia real en los sujetos de representaciones mentales de nivel subconceptual. Pero, entonces, el representacionalismo perdería gran parte de sus mayores atractivos. Es típico de la mayor parte de los representacionismos postular estados mentales representacionales tanto de nivel conceptual como subconceptual. Y el sostener que hay importantes generalizaciones sobre nuestra vida mental que, al no resultar claro cómo podrían ser formuladas mediante un vocabulario físico, se perderían si no postuláramos representaciones mentales de nivel conceptual que constituyeran los dominios de esas generalizaciones.

El computacionalismo clásico, por ejemplo, tiende a postular en nosotros un sistema representacional capaz de incluir como representaciones primitivas al menos todas las que podamos describir con ayuda de nuestros lenguajes naturales.<sup>1</sup> Si todas nuestras representaciones mentales se desarrollaran a un nivel subconceptual, o si fueran reducibles a ellas, tal postulación sería innecesaria. La constatación de que hay numerosas representaciones mentales que no siendo claramente subconceptuales dan pie a importantes generalizaciones es el principal argumento a favor de esa gran riqueza primitiva de nuestro sistema representacional interno.

Pero, ésta no es toda la verdad. Es cierto que muchas representaciones mentales desarrolladas a un nivel conceptual parecen irreducibles, pero también es plausible suponer que las representaciones mentales desarrolladas a un nivel subconceptual resulten imprescindibles para la constitución filogenética y ontogenética de casi cualquier sistema representacional. Sin embargo, la distinción entre los dos niveles de representación mental que hemos establecido es, en sí misma, neutral con respecto al problema de si finalmente cabe o no reducir toda representación mental a representaciones mentales desarrolladas a un nivel subconceptual. Y es bueno poder seguir manteniendo esa importante distinción con independencia del problema, tan discutido en otras épocas filosóficas, de la reducción de toda representación a algún tipo de representaciones subconceptuales.

No es decisiva aquí la reducción. Ni, tampoco, como se desprende de

<sup>1</sup> Véase, por ejemplo, Fodor (1975). Aunque esté relacionada con ella, esta tesis se mantendría con bastante independencia de la tesis del innatismo.

nuestros ejemplos, el que las representaciones mentales desarrolladas a nivel subconceptual incluyan o no referencias a determinadas sensaciones o cualidades sensoriales. Hasta cierto punto, sobre todo una vez que abandonamos toda exaltación reductiva, es posible describir en ampliaciones razonables del lenguaje de la física gran parte de esas sensaciones. Podemos, así, meter en el mismo saco de lo subconceptual tanto representaciones de objetos físicos teniendo determinadas propiedades físicas objetivas como representaciones de objetos físicos produciendo determinadas sensaciones subjetivas a través de ciertas relaciones físicas objetivas.

#### 4. Relevancia cognitiva

Las representaciones mentales también pueden tener o no una relevancia cognitiva. Y la tendrán en la medida en que desempeñen un papel causal determinante en la etiología de ciertos tipos de comportamientos que damos en llamar conductas intencionales o, simplemente, acciones. Lo importante para la relevancia cognitiva es la capacidad que algunas representaciones mentales tienen de ser parte integrante de la causa de ciertas acciones.

Al caracterizar así la relevancia cognitiva de *algunas* representaciones mentales, recojo una idea muy presente, a pesar de estar asociada con ciertas ambigüedades, en toda la literatura computacional clásica. Por ejemplo, frente a la distinción basada en la complejidad que traza Smolensky (1988) entre los sistemas dinámicos en general y los sistemas dinámicos cognitivos, Fodor y Pylyshyn (1988, nota 4) señalan que lo que distingue lo cognitivamente relevante es, más bien, el *papel causal* que ciertas representaciones tienen en la etiología de la conducta de la gente. Un río es un sistema dinámico muy complejo, y muchos de sus estados pueden tener un gran y complejo contenido representacional, pero un río no es un sistema dinámico cognitivo porque esas representaciones no juegan ningún papel causal relevante en el comportamiento del río. Es cierto que a veces se identifica directamente representacionalismo con relevancia cognitiva. Y aquí radica una de las ambigüedades a las que me refería antes. Los mismos Fodor y Pylyshyn (1988, p. 9), por ejemplo, una página más atrás del anterior texto dicen:

Since Classicists and Connectionists are both Representationalists, for them any level at which states of the system are taken to encode properties of the world counts as a cognitive level; and no other levels do.

Pero es obvio que el representacionalismo, al menos ese representaciona-

lismo mínimo que hemos descripto más arriba, no tiene por sí mismo consecuencias con respecto a la relevancia cognitiva. Si lo decisivo para la relevancia cognitiva no fuera la etiología de cierto tipo de comportamientos, podríamos claramente detectar en nuestro río del ejemplo anterior un nivel cognitivo.

Evaluar la relevancia cognitiva de una representación mental entrañaría, por lo tanto, considerar fáctica y contrafácticamente que ciertas conductas intencionales no son posibles a menos que se den ciertas representaciones mentales; representaciones mentales, en principio, desarrolladas a nivel conceptual o subconceptual y con una estructura que podría ser tanto simbólica como no simbólica.

## **5. Al fin las estructuras**

Las representaciones mentales, por último, podrían tener una estructura característica simbólica del tipo general propuesto por el computacionalismo clásico, una estructura característica no simbólica del tipo general propuesto por el conexionismo no eliminativista, otras diferentes estructuras características o ninguna estructura característica en especial. Es obvio que, tal como hemos definido estructura característica, las tres primeras opciones son mutuamente excluyentes y se aplicarían a la totalidad de las representaciones mentales. No así la cuarta. Analicemos más detalladamente las dos primeras posibilidades.

El computacionalismo clásico sostiene que las representaciones mentales sólo pueden tener una estructura simbólica y el conexionismo no eliminativista que no es necesario suponer en ellas ninguna estructura simbólica.

Tanto para el computacionalismo clásico como para el conexionismo no eliminativista, pensar es computar. Todos los procesos mentales en los que intervienen representaciones deben ser tratados como procesos computacionales. Tales procesos computacionales se definen sobre ciertas entidades psicológicamente reales identificables con representaciones con cierto contenido semántico.

El computacionalismo clásico y el conexionismo no eliminativista divergen, sin embargo, en la clase de computación que aquí ha de ser considerada relevante. Y la diferente estructura que postulan en las representaciones mentales depende de esa distinta clase de computación asumida.

### **5.1. Computacionalismo clásico**

Según el computacionalismo clásico, la clase de computación relevante



para el análisis de las representaciones mentales requiere operaciones computacionales definidas sobre símbolos (véanse Fodor, 1976; Newell, 1980; Pylyshyn, 1980 y 1984). Las representaciones mentales deben tener realmente una estructura simbólica. Las representaciones mentales deben ser símbolos, más o menos complejos, regulados por una sintaxis combinatoria y dotados de una semántica composicional.

Su sintaxis combinatoria incluiría unas reglas de formación y unas reglas de transformación.

Las reglas de formación especifican la clase de todas las representaciones mentales posibles bien formadas. Hacen esto, fijando una lista finita de representaciones primitivas y de posibles estructuras sintácticas y estableciendo que cualquier representación mental ha de ser una representación primitiva o una representación compleja. En este último caso, una serie de constituyentes sintácticos, que pueden ser a su vez representaciones primitivas o complejas, tendrá una de esas posibles estructuras sintácticas.

Las reglas de transformación especifican cómo unas representaciones mentales pueden transformarse o convertirse en otras. Y hacen esto estableciendo una lista finita de transformaciones posibles entre representaciones mentales, primitivas o complejas.<sup>2</sup>

<sup>2</sup> *Varios matices:* 1) Prefiero hablar de "reglas de formación" en lugar de hablar de "reglas de construcción" por una razón. La construcción de representaciones complejas a partir de representaciones más simples es un caso de transformación de unas representaciones en otras. Aunque no toda transformación sea una construcción, toda construcción sí que es una transformación. Identificar "buena formación" con "construcción" obliga a considerar las reglas de formación como una clase especial de reglas de transformación. Pero es obvio que las representaciones mentales no tienen por qué surgir sólo a través de transformaciones de otras representaciones ya existentes. Pueden originarse, por ejemplo, a través de procesos perceptivos, autoestimulativos, etc. Es conveniente, por tanto, distinguir la "buena formación" de la "construcción". 2) Lo que estamos diciendo acerca de la sintaxis combinatoria y la semántica composicional de las representaciones mentales podría ser dicho acerca de la sintaxis y la semántica de cualquier otro sistema representacional simbólico. 3) Debería distinguirse entre representaciones primitivas, representaciones simples y representaciones innatas. Son conceptos diferentes. Más concretamente, cuando al presentar las reglas de formación hablamos de representaciones primitivas y representaciones complejas, no asumimos en las representaciones primitivas ninguna otra simplicidad aparte de la que se deriva de no ser representaciones complejas desde el punto de vista de esas reglas de formación. Pero podrían ser complejas desde otros puntos de vista (por ejemplo, desde un punto de vista conexionista) y, por supuesto, podrían no ser innatas.

La semántica composicional de las representaciones mentales, por otro lado, haría que el significado de cualquier representación mental dependiera del significado de sus constituyentes sintácticos y de su sintaxis. El significado de cualquier representación mental depende, en último término, de su sintaxis y del significado de las representaciones primitivas.

Desde el punto de vista del computacionalismo clásico, la sintaxis permitiría encontrar un lugar natural para la semántica de nuestras representaciones mentales a través de la causalidad. La sintaxis ofrecería un puente perfecto entre la causalidad y la semántica. Si la causalidad es sensible y respeta una sintaxis, y esa sintaxis es sensible y respeta una semántica, entonces aquella causalidad será también sensible y respetará esta semántica.

La estructura simbólica de las representaciones mentales permitiría explicar ciertas características psicológicas nucleares: características como la gran sensibilidad de los procesos mentales a las propiedades semánticas de las representaciones, la productividad de nuestro sistema representacional, la eficacia causal del contenido, la sistematicidad de nuestros pensamientos, etc. Y ofrecería, además, un campo muy diferenciado para el desarrollo de teorías específicas sobre lo mental; teorías hasta cierto punto autónomas respecto de otras teorías más básicas sobre los procesos subyacentes.

Una vez establecida una estructura simbólica apropiada para nuestras representaciones mentales, bastaría definir los procesos mentales sobre algunas de esas propiedades estructurales para conseguir explicar las características que acabamos de señalar. Los procesos mentales se llevarían a cabo en virtud de la peculiar "forma" de las representaciones mentales, en virtud de que las representaciones mentales satisfacen ciertas descripciones estructurales simbólicas.

Utilizando una metáfora del propio Fodor (Fodor, 1987, p. 137), podríamos decir que, desde el punto de vista del computacionalismo clásico, si las representaciones mentales no fueran realmente algo parecido a las "sentencias" de un lenguaje, serían simplemente como "rocas" en las cabezas de los sujetos; como "rocas", eso sí, que mantienen ciertas relaciones causales y que más o menos misteriosamente son portadoras de ciertos contenidos semánticos. Y esta desnuda conectividad causal no sería, para dicha perspectiva, suficiente a fin de explicar esos rasgos psicológicos nucleares que mencionábamos antes.

## **5.2. El conexionismo no eliminativista**

El conexionismo no eliminativista propone otra forma de computación que no requiere operaciones computacionales definidas sobre símbolos. Rechaza la

disyuntiva “rocas” o “sentencias” y propone otra forma de entender la estructura característica de las representaciones mentales. Las representaciones mentales no son “sentencias” en las cabezas de los sujetos. Pero tampoco son simplemente como “rocas”. Son objetos estructurados de cierta peculiar manera no simbólica.

Una representación mental, para el conexionismo, es cierto estado particular de un particular tipo de sistemas dinámicos que llamaremos sistemas cognitivos. Los sistemas dinámicos son sistemas cuyo comportamiento, el comportamiento de las magnitudes que los caracterizan, puede ser analizado mediante sistemas de ecuaciones diferenciales derivables respecto del tiempo. De acuerdo a las precisiones que hicimos antes sobre la relevancia cognitiva, podemos decir ahora que los sistemas cognitivos serían una variedad de sistemas dinámicos en los que algunas de las magnitudes que los caracterizan tienen un contenido representacional capaz de llegar a ser cognitivamente relevante.

Las magnitudes involucradas en un sistema cognitivo muestran cómo cierta información sobre el mundo puede ser representada, procesada, transmitida y, por último, almacenada, llevándose a cabo procesos de percepción y reconocimiento de objetos del entorno, procesos de aprendizaje y uso de un lenguaje, procesos inferenciales, procesos de control motor y conducta inteligente, etc.

Los sistemas cognitivos estudiados por el conexionismo tienen una estructura básica muy simple. Y, hasta cierto punto, pueden ser explicados y presentados prescindiendo del formalismo de las ecuaciones diferenciales. Son sistemas representables mediante redes conexionistas.

Una red conexionista (curiosamente, más conocidas como “redes neuronales” fuera del ámbito estricto de la psicología) está compuesta por cierto número de unidades interconectadas. Estas unidades pueden ser de tres tipos:

- 1) Unidades de entrada, cuyos estados de activación representan determinado “input” inicial.
- 2) Unidades de salida, cuyos estados de activación constituyen el “output” representacional de la red después de un proceso de transmisión y procesamiento.
- 3) Unidades intermedias (“hidden units”), que pueden intervenir en ese proceso de propagación de la activación (procesamiento y transmisión) como etapas intermedias.

Cada unidad recibe a través de sus conexiones de entrada ciertas activacio-

nes con unos valores numéricos positivos o negativos, excitadores o inhibidores respectivamente. En función de esas activaciones, cambia su estado de activación. Muchas veces, tal función sólo depende de cierto "umbral". Finalmente, transmite ese estado de activación a través de sus conexiones de salida. Las conexiones de entrada y salida modulan la actividad que se propaga en función de un valor numérico, positivo o negativo, llamado su "peso". El valor absoluto de un peso depende de la diferencia existente entre la activación que inicialmente llega a una conexión desde cierta unidad y la activación que esa conexión consigue propagar a otra unidad con la que esté conectada. La función que se establece entre los estados de activación de cualesquiera dos unidades es creciente si el peso de la conexión entre ellas tiene un valor positivo y decreciente si ese peso tiene un valor negativo. Y la "pendiente" de esta función en cada instante (su derivada en cada punto) está determinada por el valor absoluto que tenga ese instante el peso de la conexión.

Como es habitual en otros campos donde se trabaja con sistemas dinámicos, podemos utilizar un símil hidrodinámico. Podemos imaginarnos una red conexionista como una red de cañerías comunicando cierto número de estanques de agua. Los estanques son las unidades y las cañerías las conexiones. El agua sigue un determinado curso, se propaga, a través de cañerías y estanques. Cada cañería tiene asociado un mecanismo compuesto por un grifo suplementario y un succionador. Si el grifo se abre, se realizan nuevos aportes de agua a través de la cañería. El succionador asociado con cada cañería está situado en el fondo del estanque al que desemboca la cañería. Si el succionador entra en funcionamiento, mermará su volumen de agua. Si el volumen de agua que llega por una cañería es mayor que el volumen de agua que su succionador desaloja, el peso de la cañería es positivo. En caso contrario es negativo. El valor absoluto del peso de una cañería depende de esa diferencia de volúmenes. Los estanques se llenan y vacían dependiendo del agua que reciben y transmiten a través de las cañerías (aunque podría haber, si quisiéramos, otras aportaciones y escapes de agua). Las cañerías de salida pueden situarse a cierto nivel en los estanques, en cierto "umbral".

Realmente, podríamos *construir* y *simular* cualquier red conexionista mediante un sistema de cañerías y estanques como el que acabamos de describir. Por supuesto, esto no quiere decir que estos sistemas de cañerías y estanques puedan ser, sin más, sistemas cognitivos. Para ello se necesitarían otras cosas. Se necesitaría que algunos de sus estados fueran representaciones con relevancia cognitiva.

No obstante, el símil hidrodinámico es interesante. Nos ayuda a ver los sistemas conexionistas como otra clase de mecanismos computacionales con un fuerte componente funcional, lo cual a veces se pasa por alto. Y nos permite,

además, entender fácilmente la noción de estados de equilibrio de una red conexionista. Si hacemos que durante cierto tiempo pase agua a través de nuestra red de cañerías y estanques, habrá ciertos estados de equilibrio de todo el sistema. Los estados de equilibrio consisten simplemente en que ciertos estanques acabarán manteniendo unos determinados niveles de agua más o menos constantes. Tales estados de equilibrio dependerán directamente de la cantidad de agua que inicialmente hayamos hecho pasar por el sistema y de los pesos que tengan las distintas cañerías. Aquí es donde se sitúa la memoria del sistema.

En general, el comportamiento dinámico de una red conexionista está determinado básicamente (prescindiendo de cosas como los "umbrales" y otros posibles añadidos) por los estados iniciales de activación de sus unidades y por los pesos de sus conexiones.

¿En qué consisten ahora, más concretamente, las representaciones para el conexionismo? Las representaciones, hemos dicho, son ciertos estados particulares de un sistema cognitivo estructurado como una red conexionista. Todo estado de este tipo consiste en tener tales o cuales unidades excitadas o inhibidas en cierto grado. Lo que se representa en las unidades es siempre, más o menos directamente, cierta información procedente del entorno. Y esta representación puede ser de dos formas. Puede ocurrir que cierta información del entorno sea representada por cierto estado de activación de una única unidad concreta de la red. Y puede ocurrir también que cierta información no sea representada por ninguna unidad concreta sino sólo a través de los estados de activación de cierto conjunto de unidades. Cada unidad concreta de ese conjunto de unidades seguirá representando cierta información. Pero habrá otra información que sólo podrá quedar representada a través de los distintos estados de activación de las unidades pertenecientes al conjunto.

Podemos imaginar este último tipo de representación como un vector en un espacio de activaciones de cierto número de unidades que, a su vez, representan informaciones del entorno. El predominio de un tipo u otro de representaciones hace que se distinga entre redes conexionistas de representación local ("localist networks") y redes conexionistas de representación distribuida ("distributed networks"). Por último, todos estos tipos de representaciones pueden estar desarrollados a un nivel conceptual o subconceptual. Y pueden ser relevantes desde un punto de vista cognitivo o no serlo.

La estructura de las representaciones mentales en el conexionismo sería, por lo tanto, *doble*. Por un lado, tendrían una estructura causal dependiente de las relaciones dinámicas de conectividad causal establecidas en el sistema del cual son estados. Por otro lado, podrían tener una estructura semántica dependiente de las relaciones semánticas establecidas con las unidades de la red.

Cuando la representación es *local*, las relaciones causales y las relaciones semánticas coinciden. El estado dinámico de la red determina lo que la representación local representa. Cuando la representación no es local sino *distribuida*, el estado dinámico de la red por sí solo no determina lo que la representación representa. Es necesario definir el conjunto de unidades sobre las que está distribuida la representación y cómo lo está. Pero, una vez que esto está fijado, el contenido semántico de la representación también lo está. Lo importante para el conexionismo es que *por ningún lado* surge la necesidad de recurrir a una *sintaxis*.

Recapitulemos lo que venimos diciendo. Para el conexionismo, toda representación mental es un estado particular de una variedad muy especial de sistemas dinámicos que son los sistemas cognitivos. Y las redes conexionistas son un buen modelo para analizar la estructura y dinámica de estos sistemas. Toda representación mental, ya sea de tipo local o distribuido, consistiría en tener tales o cuales unidades excitadas o inhibidas en cierto grado. Los estados de un sistema cognitivo son estados capaces de tener un contenido semántico. Tal contenido semántico resulta ser una función, más o menos compleja, de lo excitados o inhibidos que estén los nodos de la red que estructura al sistema. Los estados de un sistema cognitivo son representaciones mentales. Y algunas de esas representaciones mentales han de tener relevancia cognitiva, han de desempeñar un papel causal en la etiología de la conducta.

Contrariamente a la opinión de Fodor, las representaciones mentales para el conexionismo no son simplemente como “rocas” en las cabezas de los sujetos, “rocas” que mantienen ciertas relaciones causales y que misteriosamente son portadoras de ciertos contenidos semánticos. *También tienen una estructura semántica* que a veces se solapa directamente con su estructura causal y otras veces no.

Podríamos ahora pensar que cuando la semántica no se solapa directamente con la causalidad, no obstante la causalidad *también* podría llegar a ser sensible y respetar a la semántica. Este problema se consideraría un problema del *mismo tipo* que el de coordinar la causalidad con la semántica en el computacionalismo clásico. En principio, nada parece impedir que al igual que la *sintaxis* orienta a la causalidad en el computacionalismo clásico, en el conexionismo sea la *propia distribución y cambios de los pesos* lo que consiga el mismo efecto modulador. Por consiguiente, la semántica también podría encontrar en el conexionismo un *lugar natural* a través de la causalidad. Y las representaciones mentales del conexionismo simplemente tendrían una estructura semántica no simbólica. Pero, como vamos a ver a continuación, las cosas no son tan sencillas.

## 6. Argumentos contra el conexionismo: la necesidad de admitir una sintaxis combinatoria en las representaciones mentales

Al margen de reconocer una serie de virtudes obvias del conexionismo (gran plasticidad con componentes estructurales muy simples, asombrosa capacidad de aprendizaje, enorme resistencia a ciertas alteraciones informacionales y físicas, plausibilidad fisiológica, etc.), el computacionalismo clásico *rechaza* que los modelos conexionistas sean buenos modelos de nuestra mente. De acuerdo al computacionalismo clásico, los modelos conexionistas no conseguirían explicar algunas *características psicológicas nucleares* como las siguientes:

- \* Sensibilidad de los procesos mentales a las propiedades semánticas de las representaciones.
- \* Productividad de nuestro sistema representacional.
- \* Eficacia causal de nuestros pensamientos en función del contenido semántico de las representaciones que intervienen en ellos.
- \* Sistemática de las representaciones.

En el computacionalismo clásico, hay una serie de explicaciones estándar de estos fenómenos. Examinémoslas brevemente.

Unos pensamientos nos conducen a otros de acuerdo a sus contenidos. En esto consiste la sensibilidad de los procesos mentales a las propiedades semánticas de las representaciones. Ya hemos indicado que cuando se dispone de una semántica composicional basada en la sintaxis, y esa sintaxis se armoniza con la causalidad, basta definir los procesos mentales en relación con la "forma" de las representaciones para conseguir que las relaciones causales que constituyen los procesos mentales sean sensibles a las propiedades semánticas de esas representaciones.

La sintaxis combinatoria implica la posibilidad de tener una semántica composicional basada en esa sintaxis y la posibilidad de definir procesos mentales sensibles a esa semántica haciendo que esos procesos sean sensibles a la estructura sintáctica.

Respecto de la *productividad* de nuestro sistema representacional, del hecho de que el número de representaciones mentales sea potencialmente infinito, nos encontramos de nuevo con que la sintaxis combinatoria permite entender esto perfectamente. El carácter recursivo de sus reglas de formación

hace que el número de representaciones mentales bien formadas carezca, en principio, de límite.

Ya se admita o no la productividad como un hecho acerca de nuestro sistema representacional, es indudable que la sintaxis combinatoria implica la posibilidad de tener un sistema representacional tremendamente productivo.

¿Qué decir de la eficacia causal de nuestros pensamientos en función del contenido semántico de las representaciones que intervienen en ellos? Hablar de esto es hablar, entre otras cosas, de la relevancia cognitiva de algunas representaciones mentales. De una relevancia cognitiva guiada por el contenido. Y hablar, también, del "curso de nuestros pensamientos". Pero, si ya hemos visto cómo los procesos mentales pueden ser sensibles al contenido, debe resultar claro cierto camino por el que un pensamiento podría llegar a tener eficacia causal en función de su contenido semántico.

Tener una sintaxis combinatoria capaz, por un lado, de servir de base a una semántica composicional y capaz, también, por otro lado, de ser físicamente ejemplificada a través de propiedades físicas causalmente eficaces implica, de algún modo, poder tener pensamientos causalmente eficaces en función de su contenido semántico.<sup>3</sup>

Vayamos a la sistematicidad. Uno de los argumentos más fuertes del computacionalismo clásico contra el conexionismo está basado en la supuesta sistematicidad de las representaciones mentales. Fodor y Pylyshyn (1988), por ejemplo, nos dicen:

cognitive capacities always exhibit certain symmetries, so that the ability to entertain a given thought implies the ability to entertain thoughts with semantically related contents. We claim that such arguments make a powerful case that mind/brain architecture is not Connectionist at the cognitive level (p. 3).

Veamos algunos ejemplos. Pensar que Marga quiere a Manolo parece implicar que ha de poder pensarse que Manolo quiere a Marga. Pensar que algo

<sup>3</sup> Esta explicación tal vez sea la más polémica de las que estamos comentando. Es cierto que la sintaxis permite armonizar la causalidad con la semántica. Ya hemos insistido en ello. El computacionalismo clásico confía en que "de algún modo" así pueda reivindicarse también la eficacia causal de lo mental en cuanto mental (en cuanto portador, entre otras cosas, de ciertos contenidos semánticos). Muchos sostienen que aquí hay un paso en falso. Que esa armonía no implica la eficacia causal de lo mental en cuanto mental. Que sólo es compatible con ella. Pero que, asimismo, es compatible con un paralelismo psicofísico, e incluso con un epifenomenalismo. En lo que sigue, sin embargo, evitaremos esta importante discusión.



es marrón y cuadrado parece implicar que ha de poder pensarse que algo es marrón. Y también parece implicar que ha de poder pensarse que algo es cuadrado, etc.

Los dos últimos casos serían ejemplos de sistematicidad que admiten una explicación idéntica a la explicación de la sensibilidad. El primer caso, sin embargo, es diferente. Pero nuevamente, una sintaxis combinatoria y una semántica composicional definida sobre ella permitirían entender cómo ciertos cambios en la estructura sintáctica asignable a ciertos constituyentes pueden ocasionar cambios de contenido semántico en nuestras representaciones mentales complejas sin que se cambien los contenidos semánticos de sus constituyentes sintácticos.

Comparemos ahora todas estas explicaciones. Hemos dicho que funcionan como argumentos del computacionalismo clásico contra el conexionismo, especialmente la última. Y tienen una línea argumental muy semejante. En todas ellas se repite una misma idea: debe haber algo en la estructura de las representaciones mentales tal que

*condición 1 (condición causal):*

sea capaz de tener cierta eficacia causal encadenándose apropiadamente con ciertas propiedades físicas, y

*condición 2 (condición semántica):*

sea capaz, también, de hacer que el contenido semántico de ciertas representaciones se relacione de ciertas formas peculiares con el contenido semántico de otras representaciones.

No mantendremos más el suspenso. Ese “algo” para el computacionalismo clásico son los constituyentes sintácticos conseguidos a través de una sintaxis combinatoria en armonía con la causalidad.

Con esos constituyentes sintácticos, podemos tener una semántica composicional y, con ella, las deseadas sensibilidad semántica, productividad, eficacia causal del contenido y sistematicidad.

## **7. Soluciones vectoriales rechazadas: sumas/restas vectoriales y productos/descomposiciones tensoriales**

Hemos señalado, contra Fodor y otros computacionalistas clásicos, que en

el conexionismo las representaciones mentales también pueden ser objetos semánticamente estructurados, que no son simplemente como “piedras” causalmente conectadas y con un misterioso contenido semántico.

Una representación mental puede ser un vector en un espacio de activaciones de cierto número de unidades que, a su vez, representan informaciones del entorno.

Algunos autores conexionistas han intentado, a partir de aquí, recuperar algo parecido a los constituyentes sintácticos, algo capaz de satisfacer nuestras condiciones anteriores 1 y 2 (la condición causal y la condición semántica). Smolensky (1987 y 1991) y Smolensky, Legendre y Miyata (1992), por ejemplo, sugieren que, sin tener que asumir ninguna estructura simbólica basada en una sintaxis, podemos tener en las representaciones conexionistas una estructura semántica composicional a través de ciertas operaciones vectoriales.

La solución que se ofrece es doble. Podríamos entender, por ejemplo, el constituyente correspondiente a la representación CAFE como el resultado de restar vectorialmente la representación TAZA SIN CAFE de la representación TAZA CON CAFE. Paralelamente, la representación TAZA CON CAFE resultaría de la suma vectorial de la representación CAFE y TAZA SIN CAFE. Esta sería una solución vectorial basada en sumas y restas vectoriales. Como señalan Fodor y McLaughlin (1990), la propuesta tiene algunos problemas de coherencia interna. Pero, sobre todo, tiene el problema de que, siguiendo esta estrategia, el constituyente correspondiente a la representación de CAFE que hemos conseguido sería diferente del constituyente correspondiente a la representación de CAFE que obtuviéramos si consideramos otras representaciones como VASO CON CAFE, JARRA CON CAFE, etc. Entre todos esos constituyentes sólo podría haber ciertos parecidos de familia. Y con unos constituyentes de este tipo, tan sensibles al contexto, no se consigue ver claro qué puede hacer que todos ellos se asocien a una misma representación, en nuestro caso a la representación CAFE, a una representación cuyo contenido semántico determine composicionalmente el contenido semántico de otras representaciones complejas.

El problema de tener una semántica composicional sin asumir una estructura simbólica en las representaciones persiste. La sensibilidad al contexto invalida el anterior intento de solución. Pero aún tenemos otra posibilidad. Esta vez basada en los productos tensoriales entre vectores.

Los *productos tensoriales* entre vectores dan lugar a nuevos vectores definidos en un nuevo espacio. Si, por ejemplo,  $u$  es un vector definido en un espacio de 3 dimensiones cuyas componentes son  $\langle 1,0,1 \rangle$ , y  $v$  es un vector definido en un espacio de 2 dimensiones que tiene como componentes  $\langle 0,1 \rangle$ , el producto tensorial  $u \otimes v$  será un nuevo vector definido en un espacio de  $3 \times 2$

dimensiones cuyas componentes resultan ser  $\langle 1 \times 0, 1 \times 1, 0 \times 0, 0 \times 1, 1 \times 0, 1 \times 1 \rangle$ , lo que es igual a  $\langle 0, 1, 0, 0, 0, 1 \rangle$ .

Pongamos un ejemplo procedente de Fodor y McLaughlin (1990). Se trata de representar de manera conexionista palabras de cuatro letras ("casa", "mesa", "moto", etc.). Nuestras unidades de entrada son de dos tipos. Unidades que representan las posiciones ordinales de las letras (primera letra de la palabra, segunda letra, etc.) y unidades que representan letras (letras que pueden llegar a ocupar esas posiciones ordinales; letras como "c", "a", "m", etc.). La representación de cada posición ordinal será un vector de activación definido sobre las primeras unidades. La representación de cada letra será un vector de activación definido sobre las segundas unidades. Hay también unas unidades intermedias representando letras ocupando ciertas posiciones ordinales (la letra "c" ocupando la primera posición, la letra "a" ocupando la segunda posición, etc.). La representación de estos hechos es un vector de activación definido sobre estas unidades intermedias. El caso es que este último vector resulta ser el producto tensorial de los otros dos vectores anteriores. Y la representación de una palabra de cuatro letras en las correspondientes unidades de salida será el vector de activación que se obtenga al sumar todos esos productos tensoriales.

Los constituyentes de nuestro ejemplo (letras ocupando determinadas posiciones) *no son ya sensibles al contexto*. Esto es muy importante. Además, cierto producto tensorial podría así representar que Marga desempeña el papel de sujeto en la representación consistente en que Marga quiere a Manolo. Y otro producto tensorial podría representar que, en esa representación, se desarrolla una acción de querer. Y otro que tal acción se dirige a Manolo. La suma vectorial de estas representaciones conduciría a una representación compleja con la misma estructura que la que el computacionalismo clásico postula en las representaciones. Pero esta vez, *sin ningún recurso a una sintaxis combinatoria*.

Fodor y McLaughlin (1990, p. 197) reconocen, hasta cierto punto, el valor de esta posibilidad conexionista:<sup>4</sup>

It is, in fact, unclear whether this sort of apparatus is adequate to represent all

<sup>4</sup> Olvidando los detalles técnicos de las operaciones vectoriales, hay un viejo deseo subyaciendo a todos estos intentos. Se trata siempre de obtener ciertas estructuras con relaciones de orden (representaciones de estructuras sintácticas, por ejemplo) a partir de otras estructuras sin relaciones de orden (por ejemplo, ciertas unidades o grupos de unidades conexionistas más o menos activadas). Y, en realidad, formalmente no sería difícil conseguirlo si recordamos la idea quineana de que es posible definir, por ejemplo,  $\langle x, y \rangle$  como  $\{x, \{x, y\}\}$ . Así, es claro que  $\langle x, y \rangle$  no sería lo mismo que  $\langle y, x \rangle$ .

the semantically relevant syntactic relations that Classical theories express by using bracketing trees with Classical constituents [...] But we do not wish to press this point. For present polemical purposes, we propose simply to assume that each Classical bracketing tree can be coded into a complex vector in such fashion that the constituents of the tree correspond in some regular way to components of the vector.

El problema, pues, no es ya conseguir una estructura composicional para las representaciones conexionistas. Sin embargo, desde el punto de vista del computacionalismo clásico, sigue existiendo un problema. Los componentes tensoriales de un vector complejo, sus constituyentes, no tendrían por qué ser vectores reales de actividad actual en el sistema. Podrían ser simplemente vectores “virtuales” o “imaginarios”. No tienen por qué existir necesariamente en el sistema conexionista unidades cuya actividad se asocie a esos constituyentes obtenidos mediante productos tensoriales y sumas vectoriales. Cabe decir que esos constituyentes pueden ser tan irreales e inactuales como los meridianos y paralelos en la geografía terrestre.

Esto tiene una importancia decisiva. Como nos dicen Fodor y McLaughlin (1990, pp.197-198):

Smolensky's main strategy is, in effect, to invite us to consider the components of tensor product and superposition vectors to be analogous to the Classical constituents of a complex symbol; hence to view them as providing a means by which connectionist architectures can capture the causal and semantic consequences of Classical constituency in mental representations. However, the components of tensor product and superposition vectors differ from Classical constituents in the following way: when a complex Classical symbol is tokened, its constituents are tokened. When a tensor product vector or superposition vector is tokened, its components are not (*except per accidens*). The implication of this difference, from the point of view of the theory of mental processes, is that whereas the Classical constituents of a complex symbol are, ipso facto, available to contribute to the causal consequences of its tokenings—in particular, they are available to provide domains for mental processes—the components of tensor product and superposition vectors can have no causal status as such. What is merely imaginary can't make things happen, to put this point in a nutshell [el énfasis que se hace de la expresión puesta entre paréntesis es mío].

En conclusión, las sumas y restas vectoriales nos ofrecerían unos constituyentes no clásicos tremendamente sensibles al contexto. Pero con ello sólo conseguimos algo que, aun siendo capaz de encadenarse de alguna manera con la causalidad satisfaciendo nuestra anterior condición causal 1, no es capaz de

satisfacer la condición semántica 2 asegurando que los contenidos semánticos de nuestras representaciones se relacionen de forma adecuada. Por otro lado, los productos tensoriales nos ofrecerían unos constituyentes no clásicos capaces ahora de satisfacer la condición semántica 2, pero no la condición causal 1. Esos constituyentes podrían ser meramente virtuales o imaginarios y, así, no tener ninguna eficacia causal asegurada.

Llegamos a un punto en el que parece imposible satisfacer a la vez las condiciones 1 y 2, la condición causal y la condición semántica, fuera del marco ofrecido por una sintaxis. Pero, ¿es realmente imposible?

Considero el problema de la sensibilidad contextual difícilmente solucionable. Seguramente parte del contenido semántico de nuestras representaciones mentales, y de nuestras expresiones lingüísticas, sí sea contextualmente muy sensible. Pero otra parte de ese contenido semántico no puede serlo. El camino correcto no parece ser éste. Sin embargo, aún queda otro camino. Y el caso es que el problema del carácter virtual o imaginario de los constituyentes obtenidos siguiendo la segunda estrategia no resulta tan definitivo como es presentado por el computacionalismo clásico. Una vez percibido con claridad el problema, también está clara su solución.

### **8. Una respuesta conexionista: recuperación de algún tipo de solución vectorial y composicionalidad sin sintaxis combinatoria**

Si el problema fuera sólo el de la posible virtualidad o sentido imaginario de esos constituyentes no clásicos, sería sólo un problema de diseño computacional conexionista. Es cierto que los constituyentes no clásicos resultantes de un análisis tensorial pueden ser sólo virtuales o imaginarios. Pero no tienen por qué serlo necesariamente. No tienen por qué ser necesariamente actuales, pero tampoco tienen por qué ser necesariamente inactuales. Y entre ambos extremos, hay muchas posibilidades.

La cuestión fundamental es que cuando no es necesario que algo sea de cierta forma, si de hecho lo es, no tiene por qué serlo simplemente por accidente (“per accidens”) como maliciosamente sugieren Fodor y McLaughlin entre paréntesis en el texto anterior.

Curiosamente, nos encontramos con algo enteramente similar si consideramos el computacionalismo clásico. Tampoco es necesario que existan propiedades físicas capaces de permitir la ejemplificación de las propiedades estructurales simbólicas de las que habla el computacionalismo clásico, propiedades físicas que además, en el caso de las representaciones mentales cognitivamente

relevantes, tengan la eficacia causal deseada en relación con la etiología de la acción intencional. Sin embargo, el computacionalismo clásico las postula. Y no cree que esto sea un asunto meramente accidental. No es necesario que existan, ni tampoco que no existan, pero el que existan permite explicar muchas cosas. Exactamente lo mismo podríamos decir de esos constituyentes no clásicos que permiten cierta composicionalidad semántica sin basarse en ninguna sintaxis combinatoria. La existencia de mecanismos de actualización de esos constituyentes no clásicos, de los productos tensoriales de los que hablábamos antes no es necesaria. Pero si existen tales mecanismos, si las teorías conexionistas los postulan, podrían también explicarse muchas cosas.

Creo conveniente detenernos algo más en esto. No hay que confundir el problema del carácter no necesario (no esencial, añadido) de la actualización de los constituyentes no clásicos, de esos productos tensoriales, con el problema del carácter no necesario (no esencial, añadido) de, por ejemplo, la sistematicidad de nuestros pensamientos en los sistemas conexionistas. Son dos cuestiones diferentes que a veces se confunden.

Efectivamente, debe haber características psicológicas más esenciales que otras. Y, si hacemos caso al computacionalismo clásico, algunas incluso pueden ser necesarias para llegar a tener pensamientos como los nuestros. También es cierto que debemos evaluar nuestros modelos cognitivos (clásicos, conexionistas, etc.) tanto en relación con lo que pueden hacer como en relación con lo que no pueden hacer. Según esto, al menos respecto de cierta clase de representaciones mentales, la sistematicidad puede ser una de esas características psicológicas muy esenciales que impiden ciertas cosas.

A veces se ha criticado al conexionismo por no hacer de la sistematicidad algo esencial y necesario en este sentido. La sistematicidad no sería un fenómeno estructural sino, como mucho, el resultado biológico de una selección natural. Tendría una explicación más diacrónica que sincrónica. Sterelny (1990, cap. 8) recoge perfectamente algunas de estas críticas. Pero la solución vectorial que hemos visto, la solución basada en los productos tensoriales, abre el camino a explicaciones también sincrónicas y estructurales de la sistematicidad. La sistematicidad en el conexionismo podría también ser vista como algo esencial y necesario dadas ciertas condiciones de base. Sin embargo, el problema de la actualización o no de los constituyentes no clásicos es distinto. Es distinto porque tampoco en el computacionalismo clásico hay nada que garantice la existencia de propiedades físicas capaces de ser tenidas por las ejemplificaciones de los constituyentes clásicos y de asegurar la eficacia causal deseada. Simplemente se postulan. Y el conexionismo, por las mismas razones, también puede postular en sus modelos la existencia de mecanismos de actualización de ciertos constituyentes no clásicos, de los productos tensoriales.

apropiados. Es sólo bajo postulaciones de este tipo, bajo esas condiciones de base, como la sistematicidad y otras características psicológicas pueden convertirse en esenciales y necesarias.

Ahora bien, si se establecieran esos mecanismos de actualización de los constituyentes no clásicos, ¿no tendríamos ya todo lo que necesitamos para satisfacer las dos condiciones 1 y 2, la condición causal y la condición semántica, que imponíamos más arriba a la estructura de las representaciones mentales?

Efectivamente, parece que sí. Conseguiríamos satisfacer plenamente, según vimos, la anterior condición semántica 2 que el computacionalismo clásico impone a las representaciones mentales. Se podrían representar estructuras sintácticas, constituyentes sintácticos como, por ejemplo, que Marga desempeña el papel de sujeto en cierta representación. Y, al tener la representación mental de un coche rojo, por ejemplo, se podría tener siempre una representación mental de un coche. Representar algo entrañaría siempre poder representar los constituyentes clásicos de la representación y lo que esos constituyentes representan.

Y esta posible solución al problema de la virtualidad también garantizaría una plena satisfacción de la condición causal 1. Al ejemplificarse al modo conexionista una representación compleja, esa ejemplificación siempre podría estar causalmente conectada con ciertas ejemplificaciones de sus constituyentes capaces de desarrollar la eficacia causal deseada. Esto último ocurrirá en otros lugares del sistema, acaso también en otros momentos. Pero no importa. Siempre que se necesite, se podrá tener esa eficacia causal.

Con esos constituyentes no clásicos, capaces de hacer lo mismo que los constituyentes clásicos, podemos tener cierta semántica composicional y, con ella, las deseadas sensibilidad semántica, productividad, sistematicidad y eficacia causal del contenido. Pero, ¿qué composicionalidad sería ésta? Por supuesto, una composicionalidad no clásica. Recientemente, van Gelder (1990 y 1993) la ha llamado composicionalidad funcional.

En el caso de las representaciones mentales, existe una composicionalidad funcional cuando hay procedimientos efectivos para (1) producir una representación mental compleja dados sus constituyentes y para (2) descomponer una representación mental compleja en sus constituyentes. La composicionalidad clásica exigiría, además, (3) que cualquier ejemplificación de una representación mental compleja contenga ejemplificaciones de sus constituyentes. Pero bastarían los dos primeros requisitos, junto a la posibilidad mencionada antes de actualizar, siempre que se precisen, los constituyentes de las representaciones mentales complejas, para tener todo lo necesario a fin de explicar la sensibilidad semántica de los procesos mentales, la productividad de nuestros

pensamientos, la eficacia causal del contenido y la sistematicidad.

Por consiguiente, podríamos tener una semántica composicional sin tener estrictamente una sintaxis combinatoria en el sentido clásico y sin tener constituyentes sintácticos clásicos. Las representaciones mentales pueden tener una estructura y tener unos constituyentes (y poder así representar estados de cosas complejos y estructurados; por ejemplo, sentencias de un lenguaje) sin que esa estructura sea simbólica y esté basada en una sintaxis combinatoria.

### **9. Redefiniendo las clases de cosas que pueden ser ejemplificaciones de las representaciones mentales: los símbolos reconstituidos**

Pero hemos llegado a una situación en la que bastaría muy poco para volver a tener las representaciones mentales con una estructura simbólica del computacionalismo clásico. Bastaría tan poco como reconocer que, aunque la ejemplificación de una representación compleja requiera que se ejemplifiquen sus constituyentes sintácticos clásicos, estas últimas ejemplificaciones no necesitan, en ningún sentido claro preciso, ser distinguibles como parte de las primeras ejemplificaciones.

El computacionalismo clásico exige no sólo que las representaciones complejas tengan constituyentes sino, como acabamos de decir al comparar la composicionalidad clásica con la composicionalidad funcional, que esos constituyentes se ejemplifiquen al ejemplificarse esas representaciones complejas.

Veamos a continuación cómo podemos volver a tener, a partir de los constituyentes no clásicos que nos ofrece el conexionismo, los constituyentes clásicos y, con ellos, los símbolos clásicos. Comencemos admitiendo que tenemos una solución al problema de la virtualidad de los constituyentes no clásicos. Admitamos, por tanto, que tenemos un mecanismo de actualización tal que, en virtud de ciertas características del vector de activación correspondiente a una representación compleja, siempre pueda entrar en el juego causal una actualización de sus constituyentes no clásicos. Asumiendo que todo símbolo clásico es un constituyente clásico y viceversa, podemos redefinir ahora la clase de ejemplificaciones de un símbolo clásico así:

#### *Los símbolos clásicos reconstituidos*

Algo pertenece a la clase de cosas que ejemplifican un símbolo clásico si y sólo si es una ejemplificación de cierto constituyente no clásico o es una ejemplificación de algo representable mediante un vector de activación cuya



descomposición tensorial produzca ese constituyente no clásico.

Al ejemplificarse al modo conexionista una representación compleja, siempre se ejemplificarán ahora sus constituyentes clásicos. Estas ejemplificaciones no serán, en muchos casos, identificables de manera clara y precisa como “parte de” las ejemplificaciones de las representaciones complejas. Pero esto no importa mucho. A fin de cuentas, tampoco en nuestros lenguajes naturales, cuando son hablados y escritos, las ejemplificaciones de los constituyentes clásicos, de ciertos símbolos clásicos, son identificables de manera clara y precisa como “parte de” las ejemplificaciones de las expresiones complejas. No hace falta poner ejemplos.

Desde un punto de vista clásico lo único importante debería ser que al ejemplificarse representaciones complejas se ejemplifiquen sus constituyentes sintácticos y que esos constituyentes tengan la apropiada eficacia causal. La relación “parte de” es una relación ambigua. Entre los constituyentes y las representaciones complejas sí ha de darse tanto una relación de inclusión como una clara y precisa relación de “parte de”. Pero entre las ejemplificaciones de los constituyentes y las ejemplificaciones de las representaciones complejas sólo tiene que darse una relación de inclusión. No es necesaria ninguna clara y precisa relación de “parte de”. Pedir más haría imposible que nuestros lenguajes naturales tuvieran una sintaxis combinatoria y una semántica composicional. Y, en gran medida, la tienen.

Con la anterior redefinición de lo que pueden ser ejemplificaciones de un símbolo clásico, volvemos a tener representaciones mentales con una estructura simbólica. Volvemos a tener una estructura caracterizable mediante una sintaxis. Volvemos a tener reconstituidos los símbolos del computacionalismo clásico. La pregunta grave e inquietante que surge a continuación es: dado el máximo grado de generalidad y de funcionalismo de las descripciones computacionales clásicas, ¿serán, entonces, los sistemas representacionales conexionistas sólo implementaciones de los sistemas representacionales clásicos?

## 10. El camino de la integración

Voy a intentar defender ahora que, a pesar del resultado que hemos alcanzado, a pesar de la anterior redefinición de los símbolos clásicos, los sistemas representacionales conexionistas (y tal vez también otros sistemas representacionales con una estructura ni clásica ni conexionista, sistemas de los que no hemos hablado aquí) no son sólo implementaciones de los sistemas representacionales clásicos.

Mi argumento va a ser muy esquemático. Pero, aunque podría desarrollarse

con muchos más detalles, considero que su plausibilidad inicial es suficientemente fuerte. El argumento tiene dos fases diferenciadas. En la primera, llamaré la atención sobre el carácter limitado de las observaciones anticonexionistas del computacionalismo clásico. Sus observaciones, las observaciones que hemos estado indicando en otros apartados, sólo son pertinentes, respecto de una clase muy particular de representaciones mentales. En la segunda fase, argumentaré que las estructuras clásicas son muy condicionales. Argumentaré que es plausible suponer que, a menos que algunas representaciones mentales tengan determinadas estructuras no clásicas, las representaciones mentales relevantes para el computacionalismo clásico no podrán tener la estructura clásica sujeta.

En una palabra, voy a argumentar que la estructura clásica de las representaciones mentales

- 1) no tiene por qué ser la estructura característica de todas las representaciones mentales, y
- 2) es una estructura tremendamente condicional.

Comencemos con la primera fase. Al principio de este trabajo, señalábamos que las representaciones mentales podían ser muy variadas. Aparte del tipo de estructura que tengan, pueden desarrollarse a un nivel conceptual o subconceptual. Y pueden tener o no una relevancia cognitiva. El nivel conceptual o subconceptual dependía de la posibilidad de reducir o no su contenido semántico a un contenido expresable en el lenguaje de la física (en alguna extensión razonable de este lenguaje). La relevancia cognitiva tenía que ver con la eficacia causal de las representaciones mentales en la etiología de ciertas conductas consideradas acciones intencionales. Pues bien, las características psicológicas que el computacionalismo clásico considera nucleares (sensibilidad, productividad, eficacia causal y sistematicidad) no parecen características necesarias, en todo caso, más que respecto de un subgrupo muy particular de representaciones mentales, respecto de las representaciones mentales desarrolladas a nivel conceptual y con una relevancia cognitiva.

No quiero extenderme en ejemplos, pero no veo por qué nuestro sistema representacional interno ha de ser productivo y sistemático respecto de representaciones mentales con contenidos semánticos desarrollados a un nivel subconceptual, contenidos semánticos consistentes, por ejemplo, en cosas como que la temperatura de la habitación esté ahora mismo subiendo.

Respecto de la productividad, ya se ha dicho mucho. La productividad de las representaciones desarrolladas a un nivel conceptual parece estar avalada por

la productividad de nuestros lenguajes. Pero incluso aquí, la productividad de las representaciones parece tener serios límites. La subordinación y la repetición de operadores epistémicos, por ejemplo, fácilmente llevan la comprensión a un techo difícil de remontar. Y esto obliga a distinguir la productividad de los lenguajes públicos de la productividad de nuestros sistemas representacionales internos. Pero, si esto ocurre con las representaciones desarrolladas a un nivel conceptual, ¿por qué suponer una productividad en las otras, en las desarrolladas a un nivel subconceptual?

Respecto de la sistematicidad, Fodor y Pylyshyn (1988, pp. 39-40), por ejemplo, nos dicen:

It is not, however, plausible that only the minds of verbal organisms are systematic. Think what it would mean for this to be the case. It would have to be quite usual to find, for example, animals capable of representing the state of affairs aRb, but incapable of representing the state of affairs bRa. [...] In consequence, such animals would be unable to learn to respond selectively to bRa situations. (So that, though you could teach the creature to choose the picture with the square larger than the triangle, you couldn't for the life of you teach it to choose the picture with the triangle larger than the square.)

It is, to be sure, an empirical question whether the cognitive capacities of intraverbal organism are often structured that way, but we're prepared to bet that they are not.

Pero en mi opinión, y a pesar de la gran liberalidad con la que se caracteriza aquí sistematicidad, Fodor y Pylyshyn están destinados a perder esta apuesta. Un animal puede ser sensible a formas cuadradas más grandes que el resto de las formas de su entorno sin tener que ser sensible a formas triangulares más grandes que el resto de las formas presentes en su entorno, ni a formas circulares más grandes que el resto de las formas del entorno, etc. Puede ser, simplemente, indiferente a cualquier forma que haya en su entorno salvo a las formas cuadradas que pueda haber cuando éstas sean más grandes que las restantes. No hay nada de extraño en esto. Y no hace falta mucha ciencia para descubrirlo. Para la mayoría de los animales no humanos, *la relevancia que ciertos aspectos del entorno tienen para un sistema cognitivo no es en general sistemática*. Así, al animal sólo se le podría enseñar a responder a situaciones donde haya un triángulo más grande que un cuadrado porque distingue situaciones donde hay un cuadrado más grande que el resto de las formas del entorno de *cualquier otra situación*.<sup>5</sup> Pero, esto *no es propiamente sistematicidad*. Y lo mismo ocurre, en

<sup>5</sup> Por supuesto, eso es una idealización. Sin embargo, los casos empíricos reales están más cerca de esta idealización que de la que suponen Fodor y Pylyshyn.

muchas ocasiones, con nuestro propio sistema cognitivo y sus criterios de relevancia. Sobre todo, en relación con las representaciones mentales desarrolladas a un nivel subconceptual.

No veo claro, pues, por qué nuestro sistema representacional debe ser productivo y sistemático respecto de todas las representaciones mentales. Y tampoco veo claro por qué todas las representaciones mentales han de tener una eficacia causal en función de su contenido, y los procesos mentales en los que intervienen deben ser sensibles a esos contenidos. Las representaciones mentales sin relevancia cognitiva tienen una eficacia causal muy heterogénea. La sensibilidad al contenido puede mezclarse con otros muchos ingredientes e, incluso, puede estar completamente ausente.

Puede que la productividad y la sistematicidad tengan más que ver con nuestros lenguajes públicos naturales que con nuestros sistemas representacionales internos. Y que sólo cuando nuestras representaciones mentales sean irreductiblemente caracterizables a través de expresiones de esos lenguajes públicos naturales, como ocurre típicamente con las representaciones desarrolladas a un nivel conceptual, su productividad y sistematicidad tiendan a proyectarse sobre nuestros sistemas representacionales internos. Y puede que la sensibilidad y la eficacia causal del contenido sólo sean necesarias cuando lo que esté en juego sea una, más o menos inmediata, conducta deliberada. En general, el "curso de nuestros pensamientos" se parece más a una libre asociación de ideas que a un razonamiento guiado por el contenido. Siempre supone un esfuerzo guiar nuestros pensamientos de manera "racional". Lo raro es hacerlo.

En cuanto nos salimos de las representaciones mentales desarrolladas a un nivel conceptual y con relevancia cognitiva, los argumentos clásicos contra el conexionismo pierden fuerza.

Entremos ya en la segunda fase de nuestro argumento. A todos nos son familiares los argumentos externalistas en filosofía del lenguaje respecto de la determinación del contenido semántico. Y también la utilización de estos argumentos externalistas en filosofía de la mente y ciencias cognitivas. Ciertas consideraciones fácticas y contrafácticas sobre las situaciones reales externas a los sujetos (su entorno lingüístico y no lingüístico) son decisivas a la hora de identificar y asignar determinados contenidos semánticos a sus pensamientos. En una Tierra Gemela sin H<sub>2</sub>O no se pueden tener pensamientos acerca del agua, los "cerebros-en-una-cubeta" tampoco pueden tener pensamientos acerca del agua, etc.

Pero, si estos planteamientos son correctos, también ciertas consideraciones fácticas y contrafácticas sobre las situaciones reales internas a los sujetos podrían ser decisivas a la hora de identificar y asignar a sus pensamientos determinados contenidos semánticos. Es más, y esto es lo importante, aunque los argumentos externalistas no fueran correctos aplicados a los entornos de los sujetos y el individualismo psicológico fuera totalmente defendible, los argu-

mentos externalistas aplicados dentro de los sujetos podrían *aun* ser correctos. Podemos imaginarnos muchas situaciones internas a los sujetos en las que se satisfaga una determinada descripción computacional clásica pero, a pesar de ello, no podemos asignar a las representaciones mentales los contenidos semánticos deseados. Las cosas pueden ir aquí mal incluso a un nivel tan básico como el de las representaciones primitivas. La semántica composicional sólo funciona bien cuando las representaciones mentales primitivas tienen los contenidos semánticos correctos. Y esto no sólo depende del entorno. También puede depender de las estructuras no simbólicas de nuestro sistema representacional, de sus estructuras conexionistas o, incluso, de otras estructuras ni clásicas ni conexionistas.

De la misma manera que imaginar o descubrir que no hay H<sub>2</sub>O en una Tierra Gemela nos fuerza a no atribuir pensamientos acerca del agua o, al menos, a ser sumamente precavidos al hacerlo, imaginar o descubrir dentro de un sujeto estructuras informacionales o causales muy diferentes de las nuestras, incluso soportes materiales muy distintos del nuestro, debe hacernos ser muy cautos a la hora de atribuir pensamientos como los nuestros. Esto no es “chauvinismo” ni “provincianismo”. Podemos poner algo en entredicho y acabar aceptándolo plenamente. Se trata únicamente de dar más importancia a ciertos detalles.

Donde resulta más clara la gran importancia de algunos pequeños detalles de este tipo es en el caso de las psicopatías cognitivas. Podemos interpretar todas las psicopatías simplemente como “cubetas” computacionales o materiales en nuestra mente/cerebro. Dando la vuelta al ejemplo de los “cerebros-en-una-cubeta”, aplicando los argumentos externalistas dentro de los sujetos, podemos decir que mientras no sea plausible suponer que no tenemos “cubetas”, computacionales o de otro tipo, en nuestra mente/cerebro, no podremos asegurar que las estructuras simbólicas clásicas consiguen describir adecuadamente nuestras representaciones mentales. La estructura simbólica que postula el computacionalismo clásico es realmente una estructura muy condicionada.

Los sistemas conexionistas, y otros sistemas representacionales ni clásicos ni conexionistas, no pueden entenderse sólo como implementaciones de los sistemas computacionales clásicos. Son algo más. Muchos de ellos podrían ser condiciones necesarias para que las representaciones mentales tuvieran una estructura simbólica clásica. Y si hay algo claro sobre las implementaciones es esto: una implementación necesaria deja de ser ya únicamente una implementación.

Lo importante es que el esquemático argumento en dos fases que acabamos de aplicar al computacionalismo clásico podríamos volver a aplicarlo a las propias estructuras conexionistas; e incluso, seguramente, a otras estructuras

que no sean ni clásicas ni conexionistas. Y, con ello, quedaría seriamente puesta en cuestión la tesis de la estructura característica que nos propusimos analizar al comienzo del presente trabajo.

En conclusión, existen buenas razones para afirmar que las representaciones mentales en general no tienen ninguna estructura característica, ninguna estructura peculiar y específica que deba ser exigida de manera esencial a todas las representaciones mentales en general. O, al menos, si algunas la tienen, no hay razones para suponer que la tienen de manera incondicionada. Más concretamente, *la estructura simbólica basada en relaciones sintácticas de algunas representaciones mentales sólo surge bajo una serie de condiciones muy especiales entre las que habría que incluir el que esas representaciones mentales también tengan otras estructuras no simbólicas.*

## 11. Balance final

\* El computacionalismo clásico y el conexionismo no eliminativista aceptan las representaciones mentales como entidades psicológicamente reales que tienen una estructura característica, una estructura esencial generalizable al conjunto total de las representaciones mentales.

\* Para el computacionalismo clásico, esa estructura es una estructura simbólica basada en una sintaxis combinatoria y en una semántica composicional definida a través de esa sintaxis. Para el conexionismo, se trata de una estructura basada en relaciones de conectividad causal y en relaciones semánticas que no requieren ninguna sintaxis.

\* En oposición a esta tesis de la estructura característica de las representaciones mentales que comparten tanto el computacionalismo clásico como el conexionismo, y que es responsable del abismo que a veces se establece entre ambos, he defendido que las representaciones mentales en general no tienen de manera esencial ninguna estructura característica.

\* Las representaciones mentales pueden ser de muchos tipos. Aparte de la estructura que se les adscriba, pueden desarrollarse a un nivel conceptual o subconceptual. Y pueden ser cognitivamente relevantes o no serlo. Estas distinciones son muy obvias, pero tienen una importancia enorme cuando nos preguntamos por la estructura de las representaciones mentales. Olvidarlas hace que parezca plausible la tesis de la estructura característica.

\* Una condición necesaria para que sea posible analizar algunas representaciones mentales, sobre todo las desarrolladas a un nivel conceptual y que tienen una relevancia cognitiva, como teniendo una estructura simbólica clásica es que esas representaciones mentales también puedan ser analizadas como teniendo otras estructuras no simbólicas.

\* No es cierto que todas las representaciones mentales deban tener esencialmente una estructura característica; una estructura, por ejemplo, simbólica clásica. Más bien ocurre lo siguiente: sólo algunas representaciones mentales tienen esa estructura clásica en una serie de condiciones no clásicas. La tesis de la estructura característica es rechazable.

\* Esto obliga a considerar los sistemas conexionistas como algo más que simples implementaciones de los sistemas clásicos. Es necesaria una integración de ambos tipos de sistemas.

\* Y es posible tal integración si tenemos en cuenta otras formas no clásicas de composicionalidad, capaces de satisfacer sin recurrir a una sintaxis las funciones semánticas asociadas a los constituyentes clásicos, y la posibilidad abierta de establecer mecanismos de actualización de los constituyentes no clásicos así obtenidos a fin de que sean capaces de asegurar la eficacia causal deseada.

## BIBLIOGRAFIA

- Fodor, J. (1976), *The Language of Thought*, Cambridge, Harvard Univ. Press.  
 — (1987), *Psychosemantics*, Cambridge, MIT Press.  
 Fodor, J. y B. McLaughlin (1990), "Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work", *Cognition*, 35.  
 Fodor, J. y Z. Pylyshyn (1988), "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition*, 28.  
 Honavar, V. y L. Uhr (1993), *Symbol Processing and Connectionist Models in Artificial Intelligence and Cognition: Steps Toward Integration*, Nueva York, Academic Press.  
 Loewer, B. y G. Rey (1991), *Meaning in Mind: Fodor and his Critics*, Oxford, Blackwell.  
 Newell, A. (1980), "Physical Symbol Systems", *Cognitive Science*, 4.  
 Pylyshyn, Z. (1980), "Cognition and Computation: Issues in the Foundations of Cognitive Science", *The Behavioral and Brain Sciences*, 3.  
 — (1984), *Computation and Cognition: Toward a Foundation for Cognitive*

*Science*, Cambridge, MIT Press.

- Smolensky, P. (1987), "The Constituent Structure of Mental States. A Reply to Fodor and Pylyshyn", *Southern Journal of Philosophy*, 26.
- (1988), "On the Proper Treatment of Connectionism", *The Behavioral and Brain Sciences*, 11.
- (1991), "Connectionism, Constituency and the Language of Thought", en Loewer & Rey (1991).
- Smolensky, P., G. Legendre e Y. Miyata (1992), "Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition", *Report CU-CS-600-92*, Computer Science Department, Univ. of Colorado.
- Sterelny, K. (1990), *The Representational Theory of Mind*, Oxford, Blackwell.
- van Gelder, T. (1990), "Compositionality: A Connectionist Variation on a Classical Theme", *Cognitive Science*, 14.
- (1993), "Beyond Symbolic: Prolegomena to a Kama-Sutra of Compositionality", en Honavar y Uhr (1993).

FACULTAD DE FILOSOFÍA - UNIVERSIDAD DE LA LAGUNA  
SANTA CRUZ DE TENERIFE - ESPAÑA

## ABSTRACT

Some of the more representative recent issues in cognitive sciences share the thesis that all mental representations must have in an essential way some determinate and specific structure. For Classical Computationalism that structure must be symbolic and it is based on relations of constituency obtained through a syntax. Non-eliminativist Connectionism rejects the appeal of a syntax. It postulates only causal relations of connectedness and, perhaps, some sort of semantical relationships other than the ones posited by Classical Computationalism. In this paper, I examine both positions defending that it is not necessary for mental representations in general to have any determinate and specific structure. Moreover, I argue that the symbolic structure that some mental representations seem to have only emerged under a set of very special conditions among which it must be included the fact that these mental representations have other non-symbolic structures. In order to show that, the acknowledgment of the great variety of mental representations and the possibility to apply some externalist arguments also inside the subjects are crucial. In the end, we are carried very far from mere implementationism. We are carried toward the need of integrated Symbolic-Connectionist-Neurological theories of higher cognition.