



UNIVERSIDAD  
COMPLUTENSE  
MADRID

**LECCIÓN INAUGURAL**  
Curso Académico 2023/2024

# Desde el factor humano hasta la inteligencia artificial: una “evolución” de lo semántico a lo sintáctico

**Fernando Maestú Unturbe**

Catedrático del Departamento de Psicología Experimental,  
Procesos Cognitivos y Logopedia  
Facultad de Psicología

Madrid, 2023



# **Desde el factor humano hasta la inteligencia artificial: una “evolución” de lo semántico a lo sintáctico**

**Fernando Maestú Unturbe**

Catedrático del Departamento de Psicología Experimental,  
Procesos Cognitivos y Logopedia  
Facultad de Psicología

**Secretaría General Universidad Complutense de Madrid**

Depósito Legal: M-19882-2023

**Corrección, edición, diseño y maquetación:**

Departamento de Comunicación e Imagen Corporativa. UCM

Impresión: Grafo. Industrias Gráficas

Quiero dar las gracias al Ilmo. Sr. Decano, Luis Enrique López Bascuas y a los profesores Javier González Marqués y Nuria Paúl por su lectura y sugerencias. También quiero agradecer a la experta en inteligencia artificial, María del Carmen Cámara, y al experto en neurociencia computacional, Gianluca Susi, por sus “afinaciones técnicas”.



Rector Magnífico, autoridades, queridos compañeros, queridos estudiantes, señoras, señores.

Este año le corresponde a la Facultad de Psicología impartir la lección inaugural del curso que comienza, y quiero que mis primeras palabras sean de agradecimiento al Decano de mi Facultad, Luis Enrique López Bascuas, que ha depositado su confianza en mí para llevar a cabo esta intervención de tanta responsabilidad. Asimismo, quiero agradecer a la Junta de Facultad que aceptara por unanimidad la propuesta del Equipo Decanal; a todos muchas gracias.

Ante un evento con tanta historia y simbolismo, es un honor, dar esta lección inaugural, en representación de los científicos que buscan las esencias del factor humano. Siendo la audiencia tan docta y heterogénea, he abordado una temática que creo nos afecta a todos, tanto en nuestro día a día, como en nuestra labor docente e investigadora.

Por tanto, haré un breve viaje “Desde el factor humano hasta la inteligencia artificial: una ‘evolución’ de lo semántico a lo sintáctico”



Sin duda estamos inmersos en un nuevo mundo donde la tecnología cada vez está más presente en nuestro día a día y condiciona gran parte de nuestros comportamientos y decisiones. Desde los sistemas que nos ayudan a solucionar problemas tomando las decisiones apropiadas (por ejemplo, para encontrar la mejor ruta que nos conduce a nuestro destino), hasta aquellos otros que, compensando nuestra frágil memoria, nos permiten recordar citas o eventos. Son muchos los ejemplos en los que la tecnología está completando y mejorando ciertas habilidades cognitivas que antes eran más necesarias en nuestra vida diaria. Este cambio está abriendo un debate social y científico acerca de si esta facilitación cognitiva nos llevará a una pérdida de capacidad o si simplemente enfocaremos nuestras capacidades en otros quehaceres cognitivos.

La experimentación en psicología cognitiva ha contribuido de manera determinante a inspirar muchos de estos sistemas artificiales o de Inteligencia Artificial (IA), ya que los descubrimientos de fenómenos psicológicos, durante décadas de investigación empírica en psicología han generado un vasto conocimiento sobre la organización de estos procesos cognitivos y sus relaciones con la arquitectura anátomo-funcional del cerebro.

Pero antes de adentrarnos en el debate de la IA y su influencia en la conducta del ser humano, conviene hacer un breve recorrido por la psicología como disciplina científica, de modo que podamos explicar cómo se organizan los procesos cognitivos y emocionales en el ser humano. De este modo, será más fácil configurar la tesis que vamos a defender en este escrito.

## **La Psicología como disciplina científica**

Uno de los principales problemas que siempre ha intrigado al ser humano es el de encontrar las causas que explican el comportamiento humano o, con otras palabras, encontrar los principios rectores de la conducta, de modo que podamos modelarla para hacerla más adaptativa. Al conjunto de conocimientos que explican las causas del comportamiento humano se le denomina habitualmente Psicología.

La idea de que la psicología debía y podía convertirse en una ciencia ganó terreno en el siglo XIX. Para hacerla realidad fueron fundamentales los avances en el campo de la filosofía, la fisiología y el desarrollo de los primeros métodos

experimentales para el estudio de la mente. Wundt (1874) consideró la psicología como la ciencia que se ocupa del estudio experimental de la conciencia y, como observarían posteriormente algunos autores del estudio, de la conducta (Watson, 1913), o del estudio de los procesos que median entre sensación y acción (ver Kellogg, 2003).

La psicología científica tiene como marco general de acción el método hipotético-deductivo. Este método se basa en la generación de hipótesis, predicciones y sistemas de comprobación de estas mediante la articulación pautada de variables dependientes e independientes conforme a un diseño experimental bien definido.

La experimentación es uno de los métodos de investigación más habitual en la psicología científica para comprobar hipótesis. Un experimento requiere la selección de unos sujetos experimentales (controlando tanto su tamaño como su carácter aleatorio), la especificación de aquello que el investigador manipula o variable independiente, y la operativización del aspecto de la conducta que se observa (y mide) o variable dependiente.

El análisis de los mecanismos internos, que median entre el estímulo y la respuesta, llevó al desarrollo de uno de los campos de investigación más apasionantes de la Psicología: estudio de los procesos cognitivos que sustentan la conducta humana. El conjunto de estudios de este ámbito, se han agrupado en una entidad globalizadora denominada Psicología Cognitiva, término que apareció por primera vez en el título de un libro publicado por Neisser en 1967 (Neisser, 1967). Siguiendo a Sternberg, (Sternberg, 2003) la psicología cognitiva estudia cómo el ser humano percibe, aprende, recuerda y piensa acerca de la información. Así, el psicólogo cognitivo puede estudiar cómo percibe la gente diferentes formas, por qué recuerdan algunos acontecimientos y olvidan otros, cómo aprenden el lenguaje, o el modo cómo piensan cuando juegan al ajedrez o resuelven un problema cotidiano (Sternberg, 2003). El estudio de estos procesos "inobservables" llevó a los psicólogos cognitivos a realizar metáforas de la mente, como un canal de información en el que ésta se trasmite de forma pasiva. También se han formulado paralelismos entre la mente y el ordenador, con origen en las concepciones cibernéticas o de la inteligencia artificial (Newell Simon, 1972), que llevan a entender la mente como un sistema activo que puede transformar la información, que controla sus acciones a partir de la información procedente del medio ambiente. Así, podemos entender la cognición como un sistema de procesamiento de la información. El objetivo de la psicología

del procesamiento de la información es el descubrimiento de la estructura y procesos internos del sistema, que son los responsables del modo en el que el sistema se comporta.

Las limitaciones de la concepción de la mente como manipuladora de símbolos abstractos mediante la aplicación de reglas formales, al igual que lo haría un software de ordenador, llevaron a algunos autores a plantear un nuevo marco de investigación que no es otro que el del cerebro. Dentro de este marco, se desarrollaron las teorías conexionistas (McClelland y Rumelhart, 1985), en las que se muestra a una mente trabajando en paralelo a través de múltiples redes neurales, que no reflejan el funcionamiento real de las células cerebrales, sino que representan modelos de los procesos cognitivos con una inspiración neural.

El objeto de la psicología desde una perspectiva básica puede conceptualizarse como el estudio de los procesos cognitivos y su necesaria descomposición en sus estructuras y subprocessos. El estudio de los procesos psicológicos básicos ha requerido de una parcelación de la cognición en diferentes áreas de interés e investigación, que en ocasiones nos han llevado a entender la cognición como un conjunto de procesos con cierta independencia funcional. Así distinguimos entre percepción, lenguaje, pensamiento, memoria, atención, funciones ejecutivas, y praxias, una división que obedece más a una aproximación pragmática que conceptual. Los problemas complejos sólo pueden ser abordados descomponiéndolos en problemas más “simples” para avanzar poco a poco en su comprensión, por ello se ha creado en torno a cada uno de los procesos cognitivos un campo de estudio específico. En realidad, la enorme interrelación entre los diferentes procesos cognitivos nos lleva a entender el procesamiento de la información como un continuo entre percepción y acción (Fuster, 1995), en el que se aprende, representa, recupera y transforma la información en un organismo pensante sujeto a las influencias de la emoción y valiéndose de un mecanismo de comunicación que es el lenguaje (Fuster, 2003). Aun así, una de las principales preguntas de la psicología cognitiva sigue siendo si podemos hablar de una teoría global de la mente.

Los investigadores en psicología, hoy en día, hacen uso de los procedimientos y diseños del método científico, de ahí que en la carrera de Psicología las asignaturas de metodología sean algunas de las más exigentes y formativas para nuestros estudiantes. Es más, la psicología ha incluido en sus modelos, además de las variables puramente conductuales, medidas fisiológicas del sistema periférico (movimientos oculares, conductancia de la piel, cardíacas, pletismografía),

o del sistema nervioso central (técnicas de neuroimagen como la Tomografía por Emisión de Positrones o PET, La Resonancia Magnética funcional o RMf, el Electroencefalograma o EEG, la Magnetoencefalografía o MEG) creando todo un cuerpo de investigación que es la neurociencia cognitiva. Recientemente los psicólogos se han aproximado a técnicas de neuromodulación por corriente eléctrica, campo magnético, ultrasonidos etc. Esto ha llevado a ampliar el enfoque y estudio de la psicología ampliando enormemente el conocimiento sobre el factor humano, además de generar una perspectiva bio-psico-social.

## **Organización de la cognición humana**

La cognición humana está formada por al menos seis sistemas fuertemente interrelacionados: memoria, atención, lenguaje, percepción, praxis y funciones ejecutivas. La psicología cognitiva ha demostrado, tanto en investigación con humanos, como en animales, que estos sistemas son multidimensionales y comportan subsistemas que, interactuando eficientemente, generarán un sistema complejo y estocástico capaz de alcanzar propiedades generativas como la “inferencia semántica”. Llamaremos a partir de aquí “inferencia semántica” a la propiedad emergente del sistema cognitivo que permite abstraernos de la realidad y por tanto posibilita la generación de nuevo conocimiento/información no directamente asociado a la realidad estimular circundante. La idea de la existencia de Dios, la teoría de la relatividad, o la creación del cubismo, serían buenos ejemplos de ello. Todavía recuerdo la primera vez que lleve a mi hijo Rodrigo al museo Reina Sofía y cuando estábamos observando un cuadro cubista me preguntó: “¿Por qué este señor pintaba tan mal?” Aquel señor era Pablo Picasso y posiblemente es imposible saber qué le llevó a generar un nuevo concepto de la pintura alejándose de lo que cognitivamente nuestra percepción construye como una realidad tridimensional. Pero sin duda, ese salto entre conocimiento previo y un esquema de pintura totalmente nuevo sólo lo puede desarrollar un sistema con la capacidad de inferencia semántica. Hablando un día con el reputado antropólogo, don Camilo José Cela Conde, comentábamos cuál era la habilidad que realmente diferenciaba a los seres humanos de los animales y su conclusión fue rotunda: ¡¡la generación artística!! Los seres humanos somos capaces de crear el elemento más bello, estéticamente hablando, aunque no tenga utilidad alguna para la supervivencia o la reproducción. Sin embargo, los animales y las plantas han evolucionado de forma que presentan una enorme belleza, pero con un objetivo pragmático. Podemos así considerar la creación artística como uno

de los mejores ejemplos de la inferencia semántica, de esa propiedad emergente del sistema complejo que es nuestra cognición.

La memoria, la atención, la percepción, la acción, el lenguaje, o las funciones ejecutivas son sistemas que se pueden llegar a formalizar en ciertas reglas sintácticas (arquitecturas cognitivas) y que incluso se podrían llegar a formular matemáticamente. Sin embargo, esto es algo que no se ha podido alcanzar totalmente con el sistema semántico y con una de sus propiedades que es la "inferencia semántica".

Estas capacidades cognitivas se han desarrollado para dar respuesta a las sensaciones internas (interocepción), a la propiocepción y a las exigencias ambientales y así ser capaces de interactuar y adaptarnos de forma eficiente a nuestro entorno. Si la interacción con el ambiente se facilitara hasta el punto en que las exigencias ambientales fueran mínimas, posiblemente disminuiría el entrenamiento de las capacidades cognitivas y nuestro sistema se adaptaría a una gozosa tranquilidad intelectual. Sabido es que el cerebro humano tiende al ahorro energético y por tanto a la economía cognitiva. Los nuevos aprendizajes generan grandes costes energéticos, como han revelado los estudios de neurociencia cognitiva con técnicas de neuroimagen funcional. Por eso es importante que el entrenamiento cognitivo se mantenga toda la vida y continuemos activos en contextos de alta exigencia cognitiva.

## **La Psicología Básica**

Tal y como indican diversos autores, la psicología cognitiva tiene como objeto de estudio la mente humana. La aparición de esta disciplina supuso la superación del paradigma conductista que pretendía estudiar sólo lo observable (estímulo y respuesta), para así centrarse en las manipulaciones y transformaciones mentales que se ejercen sobre el estímulo y que determinan la respuesta. Por tanto, la conducta no es el objeto principal, lo serán las representaciones mentales y los procesos que operan sobre ellas. El sistema cognitivo se puede concebir como un sistema complejo, de múltiples cómputos, de procesamiento simbólico o de asociaciones dinámicas. Así y tal y como señalaron algunos de los pioneros en el estudio de la cognición, se analiza el funcionamiento de los sistemas naturales de computación sobre símbolos, o los sistemas de procesamiento de la información (Johnson-Laird, 1999). Los modelos simbólicos, los que siguen la metáfora

del ordenador o los modelos de procesamiento como los conexionistas, fueron algunos de los propuestos para explicar el sistema cognitivo. Sin embargo, si atendemos a lo que planteó certeramente el profesor Ángel Riviére (1991), un modelo de la mente debe cumplir las siguientes características del funcionamiento del sistema cognitivo: 1) estructurado y flexible; 2) resiliencia; 3) procesamiento en paralelo; 4) asignación rápida de significado; 5) capacidad de completar patrones; 6) diversidad de vías de acceso y memoria; 7) adaptación. No todos los modelos de la mente, mencionados anteriormente, pueden satisfacer fácilmente estas necesidades, por lo que el problema de la explicación de la mente no era simple y había que intentar entenderlo parcialmente (por partes), para después abordar su comprensión total. El profesor Shallice, en su libro *"The Organization of the Mind"* indicaba que los modelos de la mente no habían podido explicar algunas de las conductas más relevantes del ser humano y que esto ha podido ser debido a la falta de información/atención sobre la organización funcional del cerebro (Shallice y Cooper, 2011).

Así, esta complejidad del sistema cognitivo nos ha obligado a descomponer el problema en partes: en estructuras y procesos básicos. La Psicología Básica estudia, utilizando el método científico, la organización intrínseca y la interdependencia de los procesos cognitivos básicos. Se trata de comprender y explicar cómo funciona el sistema cognitivo, cuáles son sus elementos, cómo están organizados y qué reglas siguen. Entendemos por procesos cognitivos básicos aquellas operaciones que utiliza el sistema cognitivo para adquirir y manipular el conocimiento. Estas operaciones conllevan: atención, percepción, memoria, lenguaje, praxis y funciones ejecutivas. Sin duda, la emoción condiciona toda operación mental y por tanto debemos siempre considerar su influencia en los estudios científicos. Esta descomposición del sistema cognitivo ha provocado el desarrollo de una ingente investigación en cada proceso básico, que ha provocado la existencia de campos de conocimiento relativamente distanciados entre sí. Tanto es así que enseñamos los procesos como si fueran independientes, por la falta de un modelo integrador. En realidad, aunque hemos sido capaces de descomponer en procesos básicos la mente humana, todavía no hemos sido capaces de conceptualizar la naturaleza de dichas estructuras o la forma en cómo computan. Aún se podría decir más, hemos descrito las partes, pero no hemos sido capaces de crear un modelo en el que se explique cómo interactúan entre sí esas partes.

Una forma de ahondar en la explicación del sistema cognitivo fue el aproximarse al entendimiento de su sustrato neurofisiológico y por tanto a establecer las

relaciones entre cerebro y mente. Un planteamiento cercano a esta idea fue el formulado por Marr (1982) quién planteaba que el sistema cognitivo se podría estudiar a tres niveles: 1) el físico (el cerebro); 2) el computacional (para qué sirve); 3) el algorítmico (cómo se lleva a cabo el proceso). Así, el estudio de las relaciones entre el cerebro y la mente dio pie al estudio de pacientes con lesiones cerebrales y, como consecuencia, al surgimiento de la neuropsicología como disciplina. Aunque la neuropsicología tomó una orientación más clínica, los datos de los pacientes con diferentes patologías neurológicas y psiquiátricas siguieron ayudando al desarrollo de las teorías cognitivas que intentaban explicar la mente humana.

El intento de ahondar en las relaciones entre el sustrato neurofisiológico y el sistema cognitivo en sujetos sin patología cerebral dio pie al surgimiento de la neurociencia cognitiva. Esta disciplina aprovechó el desarrollo de las nuevas metodologías de neurofisiología y neuroimagen funcional para evaluar la actividad cerebral asociada a la realización de tareas cognitivas. Esta aproximación evitaba el clásico problema de trabajar con lesiones cerebrales, que siempre conllevaba un enorme salto para explicar el funcionamiento normal de la mente humana y por tanto el desarrollo de modelos cognitivos.

De esta manera podemos ver cómo los procesos cognitivos básicos y el sistema cognitivo siguen siendo la base de estudio sobre los que giran disciplinas como la neuropsicología y la neurociencia cognitiva y, por tanto, están bien enmarcadas dentro del ámbito de la psicología básica. Basándonos en los principios de la psicología básica, por tanto, se suele explicar en las asignaturas de neuropsicología la alteración de estos procesos cognitivos como consecuencia de un daño cerebral o de un desorden psiquiátrico. Los modelos y planteamientos de la psicología básica serán los que nos permitirán interpretar, evaluar y rehabilitar las alteraciones cognitivas.

## **Un ejemplo para entender la multidimensionalidad cognitiva: la memoria**

Es probable que este proceso cognitivo básico sea uno de los más interesantes, para profundizar en él por su relevancia y complejidad dentro del sistema cognitivo humano. Aunque habitualmente se tienen concepciones erróneas acerca

de lo que es la memoria e incluso se minimiza el valor de este proceso cognitivo, frente a otros procesos e incluso con respecto a la “inteligencia”, no nos podemos imaginar cómo funcionaría nuestro sistema cognitivo sin memoria, sin representación, sin evocar la semántica de los estímulos psicológicos y físicos que se nos plantean.

La definición de la memoria puede llevarse a cabo por su función, sus características, sus sistemas, sus bases neurales, o su patología. De acuerdo con Tulving y Craik (2000), “habitualmente se cree que la memoria es la habilidad para recuperar los eventos pasados y traer a la mente hechos e ideas. La memoria y el lenguaje tienen estas funciones, pero una adecuada definición debe contener otros aspectos”. Siguiendo a estos autores, “una adecuada definición de memoria debe involucrar un entendimiento de los mecanismos cerebrales de adquisición, almacenamiento y recuperación. Estudios científicos recientes manejan la memoria a los niveles de la experiencia, la conducta, y los mecanismos neurales; cada nivel puede ser entendido en sus propios términos, pero cualquier teoría última debe también mostrar cuán diferentes son los niveles entre sí”. Como proceso, la memoria se refiere a los mecanismos dinámicos asociados con la retención y la recuperación de información de la experiencia pasada. Específicamente, los psicólogos cognitivos han identificado tres operaciones comunes de la memoria: la codificación, almacenamiento y recuperación. Cada operación, representa una fase en el procesamiento de la memoria: en la codificación transformamos la información sensorial en una representación mental; en el almacenamiento se mantiene la información codificada en la memoria; y en la recuperación se utiliza la información almacenada en la memoria (Sternberg, 2003).

Aunque realizar una definición completa de la memoria resulte complejo, diversos autores resaltan lo que esta función cognitiva supone para nosotros. “La vida de las personas tiene significado solamente gracias a la memoria. (...). Nuestro pasado inmediato y distante define quienes somos, en qué creemos, qué hacemos y qué sentimos” (Kellogg, 2003). “Gracias a la memoria somos lo que somos y sabemos quiénes somos, y nuestra vida cobra sentido de continuidad. Sin memoria, cada día, cada hora, cada instante, significarían el fin de una cosa y el comienzo de otra” (Ruiz-Vargas, 2023). “Nuestras memorias reflejan la acumulación de nuestra experiencia vital y, en este sentido, nuestras memorias son quienes somos nosotros. (...) El análisis de la memoria es una búsqueda por entendernos a nosotros mismos, una aventura que promete revelar los secretos más propios de cómo nosotros nos convertimos en lo que somos” (Eichenbaum, 2002).

Una de las perspectivas más habituales en el estudio de la memoria es el de las propiedades químicas o de los procesos fisio-morfológicos más básicos (la sinapsis, el crecimiento axónico), llegando a plantear que el fin del estudio de la memoria es describir los sucesos fisiológicos que subyacen a este proceso cognitivo básico. Huyendo de este planteamiento reduccionista, Baddeley (Baddeley, 1999) resalta que la memoria puede ser investigada a diferentes niveles de análisis: bioquímico, biofísico, fisiológico, psicológico, etc. Parafraseando un ejemplo de este autor, “si yo quiero investigar acerca de la catedral de San Paul, puedo intentar conocer la estructura atómica de las piedras que la componen, lo que llevado a la memoria es como resaltar el estudio de las particularidades de las moléculas químicas que influyen en la memoria”. Siguiendo con este autor, “el estudio de ese nivel de análisis será relevante, aunque uno puede conocer la organización atómica de las piedras de la catedral y seguir sin conocer algo de interés acerca de la catedral, por otro lado, se pueden conocer grandes cosas de la catedral sin saber nada acerca de las propiedades físico-químicas de las piedras de la catedral”.

Según Eichenbaum (Eichenbaum, 2002) la memoria puede ser investigada desde cuatro perspectivas que él denominó “las cuatro Cs”: las conexiones, la cognición, la compartimentalización y la consolidación. Las *conexiones* se refieren al nivel más básico de análisis, es decir, el de los circuitos del cerebro. Este nivel de análisis ha llevado a la conclusión de que la memoria se codifica en términos de la plasticidad de las conexiones entre las células nerviosas, modificando su fuerza o fiabilidad de comunicación a través de las sinapsis. La *cognición* se refiere a la naturaleza de las memorias al nivel más alto de análisis, el nivel psicológico. La *compartimentalización* se refiere a la cuestión de la localización de la memoria: si existen regiones específicas o si la memoria se encuentra distribuida por el córtex o incluso por todo el cerebro. La conclusión más ampliamente aceptada desde este punto de vista es que la memoria, como un todo, se encuentra distribuida, pero, al mismo tiempo, diferentes tipos de memoria se representan en otros módulos cerebrales, circuitos, vías o sistemas. La *consolidación* se refiere a cuándo y cómo las memorias se convierten en permanentes. Se ha sugerido que la memoria es, al inicio, débil y que progresivamente se convierte en resistente a la pérdida, produciéndose un proceso de consolidación mediante el que toma un carácter permanente. Aquí, nos centraremos esencialmente en el análisis del nivel psicológico, aunque también valoraremos otras perspectivas que nos ayuden a entender la organización de la memoria a través del estudio de las relaciones entre cerebro y conducta.

Hasta aquí hemos observado diferentes perspectivas y conceptos sobre la memoria, pero el conocimiento que proviene de las diferentes perspectivas en el estudio de la memoria debe encontrar un punto de encuentro. De hecho, existen psicólogos que estudian las bases fisiológicas de la memoria, lo que podría ser un ámbito más relacionado con la psicobiología, así como psicólogos que estudian la organización de los procesos cognitivos mnésicos, ámbito más relacionado con la psicología básica. Sin duda estas dos perspectivas deben tener un punto de encuentro para lograr un concepto más global y organizado de qué es la memoria. "La metodología en el estudio de la memoria debe ser esencialmente interdisciplinar. Sin una clara unión entre las diferentes disciplinas, ninguno de nuestros tres problemas más importantes -cómo se forma la memoria, dónde se almacena, y cómo se recupera- puede ser solucionado" (Fuster, 1995).

Todo lo expuesto hasta aquí nos lleva a pensar lo complejo del estudio de la memoria y, por tanto, cualquier definición que excluya alguna de las perspectivas o niveles de análisis descritos anteriormente estaría obviando una parte importante del concepto. Aquí, nos centraremos en describir los procesos cognitivos que integran la memoria, así como algunas de las perspectivas -neuropsicología- que nos pueden servir de unión entre procesos y cerebro. De esta manera describiremos esos procesos, así como sus sistemas, su organización y la posible medida de la actividad neural que los sustenta.

Santiago Ramón y Cajal (1923) fue probablemente el primero en proponer formalmente los fundamentos de la conectividad en la memoria. Según Ramón y Cajal, la creación de nuevos apéndices celulares mejoraban los contactos entre las células y organizaba nuevas relaciones entre neuronas previamente no conectadas, sirviendo como base para nuevas habilidades motoras. Ramón y Cajal estaba realmente hablando de memorias motoras que tenían su base en la sinapsis. Siguiendo la tradición creada por Ramón y Cajal, el psicólogo canadiense Donald O. Hebb (1949) formuló el principio de la formación de la memoria por la facilitación de los contactos entre neuronas; "cuando un axón de una célula A está suficientemente cerca para excitar a una célula B y de forma repetida o persistente la estimula, algún tipo de cambio metabólico ocurre en ambas células, de tal manera que la eficiencia de A, como una de las células que dispara a B, incrementa". De esta manera ocurre una coincidencia pre y post sináptica. Los estudios de Kandel confirmaron esta hipótesis, indicando que los cambios en la fuerza de las sinapsis preexistentes ocurren durante el aprendizaje de los moluscos. Estos cambios celulares en el ámbito de la sinapsis ocurren como resultado del uso, y esto sin duda influye en los conceptos de aprendizaje

y olvido en el campo de estudio de la memoria. Siendo Hebb consciente del descubrimiento de Lorente de Nó (Lorente de N ó, 1938) acerca de que diferentes axones terminan unos cerca de los otros en la superficie de la neurona, desarrolló el concepto de sincronidad y contigüidad, base del aprendizaje asociativo. Así, “dos células o sistemas de células que estén repetidamente activas al mismo tiempo tenderán a asociarse, de tal manera que la actividad en una facilitará la actividad en la otra” (Hebb, 1949). La coactivación de dos células (sensoriales) en conexión con una tercera (motora) determinará los cambios de activación que se producen en esta última. De modo que, si dos células disparan repetidamente al mismo tiempo quedarán asociadas. Esta idea es lo que tradicionalmente se conoce como segunda ley de Hebb, enfatizando la importancia de la asociación sincronizada de las células, para el entendimiento y la formación de la memoria. Estos procesos los podemos denominar sincronía convergente presináptica. Así, cuando un grupo de células convergen y de forma conjunta excitan una célula se convertirán en un grupo asociado que trabaja cooperativamente. Estos razonamientos fisiológicos estarían en consonancia con la tradición asociacionista en psicología y podemos observarlos como indicios del conexionismo.

Hasta los años 50 el conductismo dominó el panorama en psicología. Tal y como indicaba Watson (1913), alejándose de los métodos de introspección, existen determinados constructos, como la memoria, que son inobservables y por tanto no pueden ser investigados de manera rigurosa, por lo que no deben ser campos de investigación científica. La llegada del neoconductismo permitió el estudio de los mediadores entre el estímulo y respuesta, lo que abrió la posibilidad de estudiar procesos como la memoria. En este sentido se rescata la tradición del aprendizaje asociativo como marco de entendimiento general del aprendizaje humano. La psicología de la Gestalt, aunque se centró en el estudio de la percepción, enfatizó el estudio del todo, “el todo es diferente que la suma de las partes”. Rechaza que el aprendizaje se realice a través de la suma de asociaciones mentales y propone que el aprendizaje se produce mediante un proceso de reestructuración mental, *insight*, cuando el sujeto es capaz de comprender globalmente un problema. Finalmente, el trabajo de Ulric Neisser (Neisser, 1967) influyó de manera determinante en la investigación posterior al aportar el concepto de la unidad básica del procesamiento cognitivo, la información. Neisser estableció analogías entre cerebro y ordenadores indicando que el flujo de información en un ordenador y en un sistema cognitivo puede ser similar, permitiendo el desarrollo de los modelos computacionales.

A partir de los años 60 el desarrollo de la psicología cognitiva en el campo de la memoria ha sido abrumador, la teoría de los niveles de procesamiento ( Craik

y Lockhart, 1972), el modelo de la memoria operativa (Baddeley y Hitch, 1974), la memoria episódica (Tulving, 1972), los modelos de redes (Collins y Quillian, 1969), el conexionismo (McClelland y Rumelhart, 1985), etc., marcaron el camino de la investigación en la memoria hasta nuestros días, en los que la neurociencia cognitiva mediante técnicas de neuroimagen está vislumbrando la organización de las redes neurales que sustentan la memoria humana.

A lo largo de la historia de la investigación sobre la memoria se ha ido proponiendo la existencia de diferentes tipos de memoria, desde los modelos estructurales (memoria sensorial, a corto plazo y largo plazo) hasta los llamados modelos procesuales (codificación, almacenamiento y recuperación). Estimulados por estudios neuropsicológicos que mostraban cómo pacientes con daño cerebral mantenían preservadas ciertas capacidades de memoria, mientras que otras aparecían totalmente devastadas, así como por demostraciones experimentales con personas normales en las que se observaban disociaciones en el rendimiento de distintas tareas, algunos investigadores propusieron la idea de que adquirimos información sobre nosotros mismos y sobre el entorno a través de experiencias que registramos, almacenamos y recuperamos por medio de diferentes sistemas de memoria. Se puede definir un sistema de memoria como un “conjunto correlacionado de procesos.” De forma más específica podríamos decir que se entiende por sistema de memoria una estructura integrada por interacciones entre mecanismos de adquisición, retención y recuperación que siguen unas determinadas reglas operativas. Ahora bien, ¿cómo podemos caracterizar el funcionamiento de los sistemas que componen la memoria? La respuesta a esta pregunta se ha abordado desde diferentes perspectivas, y esto ha hecho que las distintas definiciones y clasificaciones descritas en la literatura tiendan a reflejar la metodología de su autor o autores. No obstante, se pueden identificar una serie de atributos o criterios generales que se han utilizado para distinguir los sistemas de memoria: 1) el tipo de información que adquieren y almacenan; 2) los principios que gobiernan su organización; 3) su capacidad de almacenamiento; 4) la duración o persistencia de la información almacenada; y 5) qué partes del cerebro son esenciales para su integridad. Basándose en estos principios y en otros conceptos acerca de qué constituye un sistema de memoria, y cómo puede distinguirse entre diferentes sistemas, Schacter y Tulving (Schacter y Tulving, 1994), en un trabajo muy influyente, identificaron cinco sistemas principales de memoria: memoria operativa, memoria semántica, memoria episódica, memoria procedimental y el sistema de representación perceptiva. Este último sistema se observaría en tareas de *priming* y podría a su vez estar compuesto por tres subsistemas: el de reconocimiento de forma de las palabras presentadas visual-

mente, el de descripciones estructurales de los objetos, y el de representación de palabras presentadas auditivamente.

El modelo de múltiples sistemas de memoria se ha contrastado con un enfoque de procesos, más centrado en las operaciones (codificación, reconocimiento, recuerdo, etc.) requeridas en tareas específicas. Actualmente parece claro que ambos enfoques son complementarios. Así, podemos considerar que el modelo de sistemas se basa en redes amplias de estructuras y sus procesos, que operan en dominios relativamente extensos; mientras que el modelo de procesos se centra en operaciones específicas que pueden funcionar en uno o más sistemas.

### **Memoria a Corto Plazo (MCP) / Memoria a Largo Plazo (MLP)**

La propuesta de una memoria primaria y una memoria secundaria fue ya enunciada por William James en 1918. Aunque esa disociación fue mantenida a lo largo del tiempo en la década de los 60, en parte gracias a los trabajos con pacientes amnésicos, se evidenciaron los sistemas neurales que sustentaban esa distinción imponiéndose en la investigación en la psicología de la memoria hasta nuestros días. En líneas generales la MCP se refiere a la retención de poca información durante un breve periodo de tiempo. En un desarrollo extendido del concepto de MCP se ha añadido la idea de Memoria Operativa, integrado por un sistema ejecutivo central que es capaz de mantener y manipular la información y poner en relación los procesos de la MCP y la MLP. Se cree que el córtex dorsolateral prefrontal juega un papel esencial en este proceso de memoria operativa. La MLP se refiere a la memoria con una capacidad casi ilimitada que mantiene la información durante largos periodos de tiempo. Muchos autores indican que la MCP refleja en realidad un “estado activado” de una porción de la MLP, por lo que es complejo realmente sustentar esta distinción incluso desde el punto de vista neurofisiológico.

### **Memoria declarativa *versus* memoria no declarativa / procedimental**

Diversos filósofos ya habían diferenciado entre conocer una cosa y saber cómo se ejecuta. No es lo mismo conocer las reglas del fútbol que saber jugar al fútbol. La memoria declarativa (saber qué) se refiere a la información que está accesible de forma consciente al conocimiento de los eventos, hechos y conceptos. De esta manera, estos conocimientos pueden ser relatados por el sujeto, declarados, de forma voluntaria. Esta información puede tener formato de proposiciones o de

imágenes. Squire en 1986 (Squire, 1986) incluye dentro de la memoria declarativa a la memoria episódica y a la memoria semántica. Parece que las regiones mediales del lóbulo temporal, el neocórtex y el diencefalo son estructuras cerebrales íntimamente relacionadas con la memoria declarativa.

La memoria procedimental o también denominada no declarativa (saber cómo) contiene las destrezas motoras. Se refiere a las habilidades y el conocimiento de los procedimientos que se elicitan a través de la acción. Es complejo declarar el modo en el que se conocen estas acciones. La memoria no-declarativa puede ser diferenciada a su vez en varios subsistemas. Así, se incluyen las habilidades y los hábitos, el condicionamiento clásico simple, el aprendizaje no asociativo (habituación y sensibilización) y el *priming* (facilitación perceptiva). Los diferentes tipos de memoria no declarativa parecen estar sustentados por distintas estructuras cerebrales. Así, el condicionamiento clásico, ejercido sobre la musculatura esquelética, estaría relacionado con el cerebelo y las respuestas emocionales condicionadas con la amígdala. El *priming* parece tener relación con el neocórtex, mientras que las habilidades y los hábitos están relacionadas con estructuras del estriado, tales como la cabeza del núcleo caudado.

Debido a la relevancia del tema que trata este documento hablaremos con algo más de detalle del modelo ACT de Anderson. En el modelo ACT-R (Anderson, 1995) la distinción entre declarativo y procedimental cobra gran interés. En el "saber algo" se puede diferenciar entre un "saber que" (saber que los pinos son árboles) y un "saber cómo" (saber cómo se monta en bicicleta). Este tipo de conocimiento parece independiente entre sí. Podemos manejar un programa de ordenador sin saber los fundamentos de su computación. El conocimiento procedimental se corresponde a las habilidades para la acción, puede poseerse parcialmente y obliga a largos procesos de aprendizaje (se puede saber esquiar, mal - regular - bien - muy bien). El conocimiento procedimental es difícilmente trasladable a sentencias verbales (sólo mediante series muy largas de proposiciones declarativas, p. ej. pasos para montar en bicicleta), mientras que el declarativo es, por definición, verbalizable. El declarativo es un conocimiento descriptivo, de hechos, que se refiere o bien a objetos o bien a acontecimientos y, al ser del tipo todo o nada (se sabe o no se sabe), su adquisición es rápida. La estructura representacional en la memoria declarativa es la red semántica donde cada unidad de representación se corresponde con un nodo. En la memoria procedimental la unidad de conocimiento es la regla de producción. El conocimiento procedimental está representado en sistemas de producciones, más que en redes semánticas. Una producción es en realidad una regla condicional, de tal

manera que la condición comprueba el estado del conocimiento del sistema, si satisface la condición, entonces se ejecuta la acción, cambiando el estado del conocimiento. Según Anderson la adquisición de conocimiento procedimental ocurre en tres fases: cognitiva, asociativa y autónoma. En la fase cognitiva, el sujeto piensa acerca de reglas específicas para implementar el procedimiento. En la segunda fase, la asociativa, el sujeto practica utilizando explícitamente las reglas, de manera extensiva y frecuente. Finalmente, en la fase autónoma, las reglas se utilizan de forma automática e implícita, con un alto grado de integración y coordinación, de forma eficaz y rápida. Al progreso en estas fases Anderson le denomina procedimentalización, el cual es el proceso general por el que transformamos información explícita lenta acerca de los procedimientos (saber qué) en una rápida, implícita, implementación de procedimientos (saber cómo). Para completar la procedimentalización son necesarios dos procesos complementarios la generalización y la discriminación. Aprendemos a generalizar, es decir a aplicar la información de situaciones específicas a un conjunto amplio de situaciones, a aplicar reglas existentes a nuevas condiciones. Además, aprendemos a discriminar, discernir información relevante de información irrelevante. En las formulaciones más avanzadas del modelo Anderson se sugiere que estos dos tipos de conocimientos, declarativo y procedimental estarían coordinados por un sistema de memoria operativa de capacidad limitada que facilita la codificación y la respuesta a los estímulos ambientales. Es quizá una de las primeras formulaciones donde se plantea que la memoria operativa es una parte activada de la MLP. Según Anderson el aprendizaje consiste en convertir el conocimiento declarativo en conocimiento procedimental. Empezamos a aprender a tocar la guitarra según nos indican las reglas fundamentales que repetimos una y otra vez hasta que convertimos el conocimiento en procedimiento. Estas ideas han tenido una gran influencia en el entrenamiento deportivo. Pero en general es una formalización de un proceso cognitivo que se pensaba no se podía investigar empíricamente, el sistema implícito.

### **Episódica versus semántica**

Tulving (Tulving, 1972, 2002) propuso una disociación entre la memoria de los episodios y la de los conocimientos. La memoria episódica se refiere a la memoria de los sucesos vividos personalmente, toda aquella información que esté ligada al espacio y al tiempo donde ocurrió (p. ej. el día de mi primera comunión). La memoria semántica se refiere a los conocimientos generales sobre el mundo, aquellos conocimientos que están desligados del espacio y del tiempo donde se aprendieron (p. ej. que la capital de Italia es Roma). En esta memoria está

contenido todo el conocimiento acerca del uso del lenguaje. Es la memoria de los conceptos. De esta manera, yo puedo pensar en qué es una bicicleta, recuperando la categoría semántica a la que pertenece, sus atributos específicos (memoria semántica) o puedo pensar en la bicicleta que me regalaron cuando tenía 7 años (memoria episódica).

Aunque Tulving ha ido modificando su propio modelo a lo largo de los años intentaremos rescatar algunos de los conceptos más influyentes en la psicología de la memoria. Desde un punto de vista evolutivo Tulving indica que la memoria episódica representaría un sistema evolucionado a partir de la memoria semántica y por tanto la primera se incluye en la segunda, pero la segunda no se incluye en la primera. Para poder poseer memoria episódica es necesario poseer memoria semántica y a su vez estas dos necesitan de la memoria procedimental. De esta manera, la memoria procedimental se asocia a la conciencia anoética (no conocimiento explícito), la memoria semántica poseería conciencia noética (consciente de las relaciones entre los eventos y objetos) y memoria episódica auto-noética (recupera vivencias personales). La conciencia auto-noética sería desde el punto de vista evolutivo la que ocuparía el rango más alto en la jerarquía al ser una capacidad muy desarrollada en el ser humano. En los niños lo último que se adquiere es la conciencia auto-noética.

Existe una disociación entre memoria episódica y semántica que proviene del estudio de pacientes con lesión cerebral. El llamado paciente K. C., tras sufrir un traumatismo craneo-encefálico, sufrió una pérdida total de su memoria episódica, pero mantuvo su memoria semántica. Otra evidencia viene del caso de Jon. Este es el caso de un niño que sufrió una lesión bilateral del hipocampo a la edad de 4 años, pero mantuvo preservado el córtex parahipocampal. La sintomatología mnésica revelaba un patrón de suma relevancia conceptual, era un paciente con inteligencia normal, y un rendimiento escolar normal, sin embargo, sufría una alteración severa de su memoria episódica, pero sin dificultades en la memoria semántica. Presentaba un rendimiento normal en tareas de reconocimiento, pero una enorme incapacidad en tareas de recuerdo libre.

Según Tulving la memoria episódica y semántica comparten diversas características como formar parte de la MLP, compartir sistemas de codificación, son dos sistemas representacionales, manejan información proposicional y puede ser expresadas simbólicamente. Con ambas se pueden establecer inferencias, se rigen por el principio de especificidad de codificación e interactúan con otros sistemas cognitivos como el lenguaje. En este mismo trabajo estos autores

reflejan la idea de que a pesar de todas estas características compartidas, estos dos sistemas presentan una clara diferenciación: la memoria semántica no depende del hipocampo, pero sí del córtex parahipocampal. Pero además la memoria episódica contiene conciencia auto-noética (conciencia de que el yo revive ahora la experiencia vivida originalmente), con capacidad de trans-temporalidad, característica no compartida con ningún otro sistema de memoria, lo que le hace diferente incluso de la memoria que denominamos declarativa, donde sólo quedaría situada la semántica. Así la memoria episódica y la memoria semántica representan dos variedades de la experiencia consciente (auto-noética y noética). La conciencia auto-noética es la que nos permite un vínculo fluido entre pasado, presente y futuro (ver Ruiz-Vargas, 2002).

### **Implícita versus explícita**

Dicotomía generada por Graf y Schacter (Graf y Schacter, 1985) que no hace referencia a dos sistemas de memoria, sino que son términos descriptivos generales que dan cobertura a distintos fenómenos (Ruiz-Vargas, 2002). La memoria implícita se define frecuentemente como la adquisición inconsciente de información estructurada. Es decir, que somos capaces de aprender información de forma consciente (explícita) e inconsciente (implícita). Estos términos hacen referencia a las instrucciones de recuperación (test implícitos y explícitos) y a los procesos de recuperación (intencional y sin conciencia / no intencional). Ya Ebbinghaus en 1885 propuso que las experiencias previas pueden ser retenidas sin la mediación de la conciencia. Es necesario distinguir la memoria implícita de la percepción subliminal, ya que no es lo mismo. En las pruebas de memoria implícita los participantes son conscientes de la existencia del estímulo, mientras que en la percepción subliminal los sujetos no son conscientes de la existencia del estímulo. Un dato de interés es que no siempre los mensajes subliminales tienen un claro efecto sobre la conducta cuando se comparan con otras condiciones experimentales.

La existencia de esta disociación de la memoria sostiene la visión de los sistemas de memoria, aún más cuando existe una evidencia clara de esta disociación en los pacientes amnésicos. Los pacientes con amnesia (frecuentemente por una lesión bilateral en las regiones mediales del lóbulo temporal) tienen serias dificultades para establecer memorias declarativas, pero por el contrario presentan un aprendizaje no-declarativo o implícito dentro de la normalidad. El mejor ejemplo lo constituye el paciente H.M que era capaz de aprender cómo trazar los bordes de una figura mirando a un espejo en vez de mirando a la figura. Quizá

otro buen ejemplo de esta disociación lo constituya un experimento en el que se observó que la modificación de las características físicas de una palabra (mayúsculas-minúsculas), reduce el efecto *priming* (facilitación perceptiva), pero no tiene ningún efecto sobre el recuerdo explícito. Por el contrario, el procesamiento semántico de las palabras mejora su recuerdo explícito, pero tiene pocos efectos sobre el *priming*.

## **Conexionismo y memoria**

Diversas teorías de la psicología de la memoria han ofrecido una visión de la memoria como una librería de almacenamiento de estímulos. Otras han propuesto que la memoria se estructura en un conjunto de nodos (Anderson, 1995) con conexiones asociativas entre ellos. El conexionismo abandona esencialmente estas visiones sobre la memoria para plantear que lo que está almacenado en la memoria es un conjunto de cambios en las instrucciones que las neuronas se envían las unas a las otras, afectando a patrones de actividad que pueden ser contruidos para un estímulo dado. De este modo, cuando se observa un estímulo se crea un patrón de actividad a través de una serie de unidades de procesamiento. Este patrón de activación será considerado la representación del evento. El patrón de representación de la activación da lugar a la creación de las instrucciones. El grupo de instrucciones es almacenado en las conexiones entre las unidades, dónde está disponible para la construcción de futuros patrones de actividad. La recuperación de la información se lleva a cabo mediante un proceso de reconstrucción de los patrones de activación de los estímulos. Así, el conexionismo se basa en la idea de que el conocimiento consiste en una serie de valores o pesos de las conexiones entre unidades simples de conexión. Estas ideas inspiraron de manera determinante los algoritmos de redes neuronales de la IA.

La psicología cognitiva ha presentado un enorme desarrollo gracias a las teorías computacionales. Tanto en el caso de los ordenadores como en el del cerebro se trata de sistemas físicos que procesan símbolos. Sin embargo, los modelos conexionistas indican que considerar al cerebro como un mero procesador de símbolos no nos puede permitir conocer la enorme complejidad de la cognición humana. Por el contrario, estos modelos asumen que el cerebro es un procesador complejo de información capaz de procesar gran cantidad de datos en paralelo, mientras que el ordenador presenta un procesamiento esencialmente serial. Así, los procesos mentales se consideran como el resultado de la transmisión de niveles de activación en grandes redes de unidades simples de procesamiento interconectadas, de forma abundante, entre sí, y no como el resultado de ex-

presiones simbólicas estructuradas, tal y como plantea las hipótesis clásicas (McClelland y Rumelhart, 1985).

## **Los nodos**

Así, surgen los modelos conexionistas que se basan en modelos que logran superar los modelos simbólicos de redes y estructuran la cognición modelando un sistema de conexiones y pesos de esas conexiones. El Procesamiento Distribuido en Paralelo (del inglés *parallel distributed processing* o PDP), el cual frecuentemente se ha llamado el modelo conexionista, indica en esencia que la representación del conocimiento reside en las conexiones entre diferentes nodos y no en cada uno de ellos, tal y como se propugnaba en las teorías clásicas de redes. Por tanto, no hay un lugar específico donde resida la memoria, cada peso está involucrado en la memoria de cada ítem. Un ejemplo del lenguaje nos permite entender mejor este concepto. Las letras no nos dicen nada es la combinación de diferentes letras y las posibles combinaciones entre ellas las que forman una palabra.

Los primeros modelos conexionistas eran semejantes a los modelos simbólicos asociativos (Anderson, 1983). En ellos los ítems complejos están representados por unidades que están conectadas con otras unidades más elementales. Sin embargo, los modelos conexionistas se componen de unidades simples, o nodos, que pueden activarse o no, y conexiones entre los nodos, estando la información representada en los pesos de las conexiones. Esta idea lleva consigo la necesidad de representaciones distribuidas a lo largo de todas las conexiones. Tal y como se formula, la activación de un nodo provocará la activación de otro nodo conectado y este proceso de difusión de la activación provocará la activación de otros nodos adicionales. Esta propagación de la activación difiere de la propuesta por Quilliam (Quillian, 1968) ya que en el conexionismo la propagación de la activación se produce en una sola dirección.

Los nodos son de tres tipos: 1) unidades de entrada que reciben información de otras redes; 2) unidades de salida que mandan la información a otras redes; 3) unidades escondidas que no reciben ninguna activación desde fuera de la red local. Cada nodo posee un umbral de activación, el cual está representado por la cantidad de energía requerida para realizar un disparo. Además de la existencia de los nodos, una de las ideas centrales de estos modelos es que los nodos están conectados entre sí, observándose conexiones en cada nodo desde cada uno de los otros nodos de la red, de aquí el nombre de conexionismo. Las conexiones

pueden ser de dos tipos: 1) excitadoras e 2) inhibitoras dependiendo de sus pesos. Así las excitadoras darán como resultado valores positivos, mientras que las inhibitoras presentarán valores negativos de salida.

## **Componentes**

Rumelhart (Rumelhart, 1989) propone siete componentes de un sistema conexionista:

1. Un conjunto de unidades de procesamiento.
2. Un estado de activación definido sobre las unidades de procesamiento.
3. Una función de salida por cada unidad.
4. Un patrón de conectividad entre unidades.
5. Una regla de activación para las entradas que inciden en una unidad.
6. Una regla de aprendizaje por la que los patrones de conectividad son modificados por la experiencia.
7. Un medio dentro del cual el sistema debe operar.

## **El aprendizaje**

El aprendizaje en un sistema de este tipo asume la idea de un comienzo como una "tábula rasa" (Aristóteles), así antes de que una red esté formada, los valores de las interconexiones están dispuestas aleatoriamente. Cada vez que se presenta un estímulo el modelo genera una salida, si es correcta, se altera el sistema de pesos para hacer más probable esa respuesta, si la respuesta es incorrecta se modulan los pesos para hacer esa conducta menos probable. Por tanto, cada vez que utilizamos el conocimiento modificamos sus patrones de representación. Lo que está almacenado no es un patrón específico de conexiones, es un patrón de fuerzas de conexiones excitatorias e inhibitorias. La nueva información refuerza o disminuye la fuerza de esas relaciones, modificando los patrones de conexión. De esta manera se genera la llamada "regla delta" que sirve como sistema de detección de errores. El aprendizaje está dirigido por la diferencia entre la señal externa entrante y la señal interna entrante a una unidad de recepción específica, generando un valor delta determinado. Así, se intentará que la entrada externa y la interna a la unidad presenten valores equilibrados de manera que las conexiones internas se vayan ajustando hasta eliminar el factor de error. Otro sistema de aprendizaje es la "retropropagación". Con este sistema la respuesta resultante a la propagación directa es comparada con la respuesta realmente correcta demandada. Así, los pesos de las conexiones entre las unidades serán reajustados cuando exista diferencia entre la salida actual y la salida correcta esperada.

El olvido podría venir justificado por el mismo mecanismo que gobierna el aprendizaje, modificando los pesos de tal manera que la salida actual sea similar a la salida deseada. El cambio de los pesos cambiará la salida de lo que se aprendió previamente.

## **La recuperación**

Cada unidad de procesamiento posee un valor de activación. El valor de activación pasa a producir un valor de salida a través de una función. El valor de salida pasa a otras unidades en el sistema. Un aspecto relevante es que existe un peso o potencia asociado con cada conexión que determina el grado de ligazón entre las unidades. Podríamos resumir los siguientes mecanismos de recuperación, (a) satisfacción de restricciones, (b) procesamiento interactivo, (c) igualación de pautas o memoria dirigida al contenido, (d) generalización automática o representación directa de la similitud. Así, podemos decir que la recuperación se alcanza por la satisfacción de un gran número de restricciones que interactúan entre ellas. Cuando la búsqueda de una información se hace compleja en una red, los sistemas que presentan un procesamiento serial mostrarían muchas dificultades para solventar el problema, sin embargo, las arquitecturas conexionistas presentan más facilidades ya que trabajan con una serie de restricciones para encontrar la información que mejor se empareja el patrón que se está buscando.

El proceso es reconstructivo, ya que lo que se almacena no es la representación exacta del objeto, sino una serie de instrucciones y pesos de conexión cuya activación produce la reconstrucción de los patrones de representación. Al estar todas las unidades de la red conectadas entre sí, la activación de una parte puede provocar la reconstrucción de las otras características interconectadas. No es necesario que la información entrante reproduzca el patrón exacto almacenado sobre un estímulo particular. De hecho, cuando accedemos a la información en situaciones de mala calidad estimular, con que algunos de los patrones distintivos sean activados será suficiente como para que el estímulo sea reconocido. Así, en situaciones de estimulación degradada nuestro sistema puede responder flexiblemente basándose en información parcial.

## **Conexionismo y cerebro**

Son modelos inspirados en el funcionamiento del cerebro, pero no intentan simular el funcionamiento específico de las neuronas, se basan más en sus patrones de interconexión formando redes complejas que sustentan el conocimiento. Un

ejemplo de paralelismo con el cerebro se encuentra en la cronometría mental desarrollada en estos modelos. Una neurona tardará alrededor de 3ms en responder a un estímulo. Si el procesamiento fuera serial sería demasiado lento como para dar una respuesta rápida al estímulo. Sabemos que podemos responder a un estímulo del medio ambiente en unos 300ms, lo que parece indicar que nuestro procesamiento es en paralelo involucrando diferentes procesos que se ejecutan al mismo tiempo para dar una respuesta integrada y rápida al estímulo. Nuevamente los modelos conexionistas asumen que el conocimiento se sustenta sobre redes neurales declarativas que se interrelacionan entre sí y funcionan en paralelo. Así, toda forma de conocimiento está representada en forma de red.

### **PDP y sistemas de memoria**

Tal y como resalta Sternberg (2003), el modelo PDP encaja elegantemente con la noción de memoria operativa comprendiéndola, como la porción activada de la memoria largo plazo (MLP). Es interesante destacar que tal y como indica la teoría del PDP, la activación se difunde a través de los nodos dentro de una red, mientras la activación no exceda los límites de la memoria operativa. Quizá una de las características más interesantes de las hipótesis conexionistas es su capacidad de integración de diferentes conceptos de la memoria. La memoria operativa trabaja a través de un procesamiento en paralelo y comprende la porción activada de la MLP; la difusión de la activación incluye la activación simultánea (en paralelo) de múltiples conexiones entre los nodos dentro de una red. Así se comprende la posibilidad de procesar de forma simultánea múltiples operaciones.

## **Las emociones**

Dada la importancia de las emociones, en el estudio de la psicología y para entender la complejidad de la mente humana, dedicaremos un apartado a explicar brevemente esta característica esencial de lo que somos.

No seríamos lo que somos sin emociones. Nos confunden, nos llevan a precipitarnos, a ser imprecisos, a cometer actos irracionales, conductas que en ausencia de un contexto o volición emocional nunca habiéramos cometido. También nos hacen enfermar. Sin embargo, las emociones nos hacen felices, nos alertan del peligro, accionan el marcador somático hacia las personas y las cosas, mejoran nuestra motivación y nuestras funciones cognitivas. Aprendemos mejor con emociones, atendemos y percibimos mejor, nos ayudan a enfrentarnos a nuestras

barreras físicas y cognitivas. Tal y como indica el profesor Luis Aguado (2019) “sea persiguiendo grandes metas vitales, buscando experiencias novedosas y excitantes, acercándonos a nuestros semejantes o procurándonos pequeños placeres cotidianos, la búsqueda del bienestar y la felicidad es la principal motivación de la conducta humana”. Así esa motivación emocional o esa disposición para la acción mueve a la conducta y al pensamiento llevándonos a extremos que difícilmente serían alcanzables sin la emoción. En esencia, las emociones nos inducen un estado fisiológico y mental que nos preparan para cualquier evento de nuestro mundo interoceptivo o del mundo estimular exterior a nuestro cuerpo.

Son interesantes los experimentos que estudian tanto las resecciones bilaterales de la amígdala, como el síndrome genético que produce su calcificación. La amígdala es una de las estructuras cerebrales más importantes del circuito límbico que rige las emociones. Klüver y Bucy llevaron a cabo unos estudios en los que realizaban una resección bilateral de la amígdala a primates no humanos. Cuando al macho dominante se le realizaba esta resección pasaba al último eslabón de la jerarquía, ya que su conducta dejaba de tener la motivación por el dominio. Los monos se volvieron desmotivados y sus expresiones faciales y vocalizaciones se hicieron menos expresivas. Parecían haber perdido la sensación de miedo a cosas que antes les hubieran dado pánico y tenían una conducta hipersexualizada. Algo semejante ocurre en el síndrome de Urbach-Wiethe en humanos. Este síndrome se produce por calcificaciones bilaterales de la amígdala debido a depósitos de calcio en los vasos sanguíneos en el interior de esta región del cerebro. Las personas que sufren esta dolencia pierden específicamente la sensación de miedo, sin aparentemente afectar a otras emociones. El miedo permite identificar contextos de peligro, y por tanto activa los impulsos de lucha o huida. Estas personas suelen ser agresivas e incluso paranoicas. Uno de los casos mejor documentados es el caso S. M. descrito por Justin Feinstein, neuropsicólogo del Instituto Tecnológico de California: “S. M. vio las películas de terror que a casi todo el mundo le ponen los pelos de punta, pero en ella no se observó ninguna reacción. El doctor le llevó a conocer especies exóticas como víboras o insectos venenosos. La mujer ni se inmutó al ver una serpiente pitón y los investigadores tuvieron que intervenir para que no acariciara una tarántula”.

Después de ver estos casos, ¿cómo podríamos haber sobrevivido como especie sin la sensación de miedo?

Las emociones tienen un importante componente fisiológico, generando importantes reacciones corporales (cardíacas, gástricas, hormonales, etc.). La excesiva

respuesta fisiológica puede bloquear el sistema cognitivo, generando conductas complejas y ciertamente irracionales. El aprendizaje del control de las respuestas emocionales ha ocupado un gran cuerpo de investigación en psicología, tanto para el tratamiento de fobias, trastornos de ansiedad, estrés. De esta manera, se han desarrollado protocolos y técnicas de intervención específicas como la desensibilización sistemática, la inundación, relajación, el control cognitivo y en las últimas décadas modelos de mindfulness y meditación. Todas ellas se han sometido a estudio mediante el método científico y se han validado en múltiples poblaciones a lo largo del planeta.

El psicólogo Paul Ekman (1983) describió 6 emociones básicas que se dan en la mayoría de las culturas: sorpresa, asco, tristeza, ira, miedo, y alegría. Aunque se han investigado otras emociones, este modelo de la emoción es uno de los más aceptados.

Además del modelo de las 6 emociones también ha surgido la teoría de la dimensionalidad de las emociones con dos importantes dimensiones: la valencia (valor positivo/placentero o negativo/aversivo) y el arousal (intensidad o nivel de excitación). Este sistema dimensional ha permitido establecer sistemas de medida de las emociones con una aproximación empírica y bien validada. Un ejemplo de esto es el International Affective Picture System (IAPS), que es una base de datos de imágenes diseñada para proporcionar un conjunto estandarizado de escenas para estudiar la emoción y la atención. El IAPS fue desarrollado por el Centro Nacional de Emoción y Atención del Instituto Nacional de Salud Mental de la Universidad de Florida, dirigido por Peter Lang, recientemente retirado. Comprende, aproximadamente, 1.000 fotografías en color que van desde objetos y escenas cotidianos, hasta escenas menos frecuentes pero emocionantes, como cuerpos mutilados y desnudos eróticos. Es un sistema validado en muchos países del mundo y ha permitido consensuar hallazgos en diferentes culturas y comparar fácilmente los resultados. Sus estímulos han sido utilizados en pruebas de neuroimagen con RMf o MEG con gran impacto en la literatura científica internacional.

La psicología en su intento de conocer de forma más específica la psique humana, no sólo desde la descripción de la patología, sino desde lo que es característico de la normalidad, ha realizado múltiples investigaciones con mayor o menor éxito. Una corriente que en las últimas décadas está ganando más importancia es la denominada "Psicología Positiva" que estudia las bases del bienestar psicológico y de las fortalezas y virtudes del ser humano. ¿Qué es lo que permite dar valor a la vida, y qué factores contribuyen para vivir una vida plena? Quizá

este enfoque tiene la característica de centrarse por primera vez en los aspectos positivos del ser humano en vez de focalizarse en sus procesos psicopatológicos, lo que claramente ha supuesto un magnífico complemento al anterior enfoque más centrado en la psicopatología. En palabras del Catedrático de Psicología de la UCM y académico de la Academia de Psicología de España, Carmelo Vázquez, “Los hallazgos de estos últimos años han comenzado a integrarse en la práctica psicológica dando lugar a numerosas intervenciones. (...) el hecho de adoptar una perspectiva positiva-complementaria a otras más tradicionales y bien asentadas dentro de la Psicología- puede ser muy fructífero para abrir nuevas fronteras en campos tan diferentes como la psicoterapia, la educación o la gestión de recursos humanos”. De hecho, desde la UCM se colidera un proyecto internacional de psicología positiva, en colaboración con Martin Seligman (U. Pennsylvania). Es un proyecto internacional para la validación en diversos idiomas de instrumentos de evaluación del bienestar psicológico y la felicidad que pueden completarse on-line ([www.psicologiapositiva.org](http://www.psicologiapositiva.org)).

Sin duda las emociones tienen una íntima relación con los procesos cognitivos. Su influencia favorece el procesamiento cognitivo, como también puede inducir su bloqueo o fracaso. Como ya hemos comentado en el apartado sobre la memoria, las experiencias emocionales generan un patrón distintivo que favorece el aprendizaje. Craik y Lockhart (1972) demostraron que incluso en ausencia de intención de aprendizaje, un procesamiento profundo (p. ej. emocional) de la información hace que aprendamos. Esto lo podemos entender bien en acontecimientos de alta intensidad emocional positivos (el día que conocí a mi mujer) o negativos (una experiencia traumática). Ambos son difíciles de olvidar, porque, aunque la experiencia sea única y puntual son de una alta intensidad emocional, facilitando el procesamiento semántico profundo y por tanto el aprendizaje. En este sentido, se ha demostrado cómo aquellos docentes que saben transmitir emocionalmente la información de sus asignaturas, obtienen mejores rendimientos en sus alumnos que aquellos que transmiten la información con menos pericia emocional. Lo mismo que ocurre con la memoria, puede ocurrir con otras funciones cognitivas donde la emoción tiene una influencia esencial, por ejemplo, en el lenguaje, atención, acción, o funciones ejecutivas.

Es en este último proceso donde el neurólogo portugués, premio Príncipe de Asturias, Antonio Damasio, desarrolló su modelo del marcador somático. Según esta teoría la toma de decisiones depende de cambios homeostáticos generados por el cuerpo. El cuerpo envía señales en forma de cambios físicos que anticipan la toma de decisiones y, sobre todo, los posibles resultados de dichas elecciones,

disminuyendo, en gran medida, la carga de trabajo en el procesamiento cognitivo racional. El marcador somático puede estar mediado por estímulos que se procesen de forma implícita, pero que generan respuestas conscientes que son difíciles de racionalizar (Damasio 1996). Al igual que en la toma de decisiones, las emociones cobrarán un papel fundamental en los procesos de control cognitivo (inhibición de la conducta), donde aquellos estímulos con una alta carga emocional positiva tenderán a generar más respuestas de aproximación, produciendo conflictos conductuales. Si algo caracteriza al ser humano frente a los animales es su capacidad, sobredimensionada, para ejercer un férreo control sobre los estímulos emocionales y de esa manera dejar que prevalezca el esquema cognitivo previamente establecido, frente al impulso inducido por el estímulo emocional. Un ejemplo sería el control de las emociones durante una acalorada discusión científica, donde los argumentos racionales deben prevalecer frente a las posibles inflexibilidades que generen las emociones.

Finalmente, es importante destacar para el tema que nos ocupa, que las emociones tienen una enorme influencia en nuestro sistema semántico e influyen de manera determinante en la generación de nuevas categorías, llegando incluso a reconstruir antiguos conceptos en función de las nuevas experiencias emocionales. Esto hace que las emociones sean plenamente partícipes en la conducta y acción humanas, construyendo y determinando nuestra mente y nuestro éxito como especie. Posiblemente, lo que indujo a Picasso a desarrollar el Cubismo fue una necesidad de superar lo establecido, una motivación extra provocada por la emoción, que llevó su capacidad cognitiva al límite para generar una nueva categoría artística. La cognición y la emoción llevan al ser humano más allá de las reglas sintácticas y le permiten crear nuevos conceptos semánticos.

## **Los nuevos modelos de neurociencia cognitiva sobre el funcionamiento de la entente cognición-cerebro: la teoría del conectoma**

En las últimas dos décadas la psicología está viviendo un renacimiento del conexionismo, corriente que se mantuvo con un carácter muy marginal desde los años 80, empujado por las evidencias de los estudios neuroanatómicos en

neurociencia. De esta forma ha surgido la teoría del conectoma. En nuestra opinión las investigaciones y fuente de datos de este modelo van a generar muchos avances en el futuro de la IA.

El nuevo modelo de conectoma permite distinguir diferentes organizaciones funcionales cerebrales como el “mundo pequeño” o las “redes de escala libre”. Las estructuras de mundo pequeño hacen referencia a la capacidad del cerebro que permite que una serie de nodos tengan una fuerte comunicación con sus vecinos, formando una comunidad, y alguno de ellos puede conectar con otros nodos a larga distancia generando una red de alta eficiencia y un equilibrado coste energético. Por otro lado, las redes de escala libre se caracterizan por la existencia de *hubs* o regiones de alta densidad de conectividad que permiten un flujo de información más eficiente, y cuyo daño genera disfunciones e importantes síndromes neuropsicológicos. Estas redes funcionales pueden ser descritas por técnicas de neuroimagen, como la magnetoencefalografía (MEG), que destaca por su utilidad para describir el espacio-tiempo-frecuencia en el cerebro, o como la resonancia magnética funcional (RMf). Esta teoría puede explicar algunas patologías neurológicas de manera eficiente, como la enfermedad de Alzheimer o el traumatismo craneoencefálico.

### **El modelo de la localización funcional**

A lo largo de la historia, el ser humano se ha preguntado frecuentemente de dónde surge su excepcional capacidad cognitiva y cuál es el sustrato fisiológico de su conducta. Así surgió progresivamente el estudio de las relaciones entre el cerebro y la conducta. Los estudios sobre las relaciones entre cerebro y conducta se realizaron durante décadas a través de las alteraciones cognitivas que presentaban pacientes con lesiones (método lesional) o mediante lesiones experimentales en animales. Con la llegada de las técnicas de neuroimagen, estas relaciones pudieron ser exploradas también en sujetos controles, lo que evitaba el siempre incómodo problema de la reorganización provocada por la lesión. Los datos que proporcionaron estos estudios confluyeron hacia el modelo de la especialización funcional, también denominado localizacionismo. Según esta hipótesis, las regiones del cerebro están especializadas en ciertos procesos cognitivos. Un ejemplo lo puede constituir el lenguaje expresivo y comprensivo. El neuroanatomista francés Paul Broca (1824-1880) pudo demostrar que las lesiones en la tercera circunvolución frontal (en la vecindad del giro triangularis) generaban dificultades de expresión del lenguaje o afasia de producción (también denominada afasia de Broca). Curiosamente estos pacientes preservaban la

capacidad de comprensión lingüística. Posteriormente el neurólogo y psiquiatra alemán Karl Wernicke (1848-1905) describió cómo otros pacientes, con lesiones en la corteza temporal superior, presentaban alteraciones en la comprensión del lenguaje, manteniendo su habilidad de producción (aunque tenían dificultades para repetir palabras). Se comprobó también que las lesiones bilaterales, en las regiones mediales del lóbulo temporal, provocaban amnesia (como en el caso H. M.), y las lesiones en regiones prefrontales causaban disfunción ejecutiva (como en el caso de Phineas Gage). Otros casos clínicos demostraron que otras lesiones a lo largo del córtex cerebral podían provocar específicamente agnosia o apraxia. Los descubrimientos del neuroanatomista alemán Korbinian Brodmann en 1909, acerca de la existencia de regiones cerebrales con cierta diferenciación citoarquitectónica, fundamentaron todavía más la especialización funcional, ya que ofrecieron el sustento fisiológico a estos hallazgos conductuales: diferenciación fisiológica coincide con disociación funcional.

Los estudios de estimulación cerebral profunda y el uso de fármacos que incrementaban la inhibición fisiológica y afectaban a la conducta, todavía reforzaron más los hallazgos previos. Este es el caso de los estudios realizados en el Instituto Neurológico de Montreal por el neurocirujano Wilder Penfield y el neurobiólogo y neurólogo canadiense Herbert Jasper, con pacientes a los que despertaban durante la cirugía y que se les estimulaba eléctricamente una parte de su córtex mientras realizaban tareas como hablar. La estimulación eléctrica sobre las regiones del lenguaje expresivo provocaba el fenómeno de “parada del lenguaje”, que no se producía cuando se estimulaban otras regiones del córtex no relacionadas con esta función cognitiva. Estas estimulaciones directas sobre el córtex cerebral evidenciaban que determinadas regiones cerebrales estaban especializadas en determinados procesos cognitivos al quedar intactos otros procesos. Un ejemplo del uso de la farmacología para evaluar funciones cognitivas fue el denominado “test de Wada” o “test de amital intracarotídeo”, desarrollado por el neurólogo japonés Juhn Wada en la década de 1950. En esta prueba se inyectaba un fármaco que inhibía temporalmente la actividad neuronal de un solo hemisferio y de esta manera se podía comprobar la lateralización hemisférica del lenguaje y realizar una simulación de las consecuencias que pudiera tener una intervención quirúrgica para tratar la epilepsia sobre la cognición. Los hallazgos obtenidos mediante los procedimientos anteriormente descritos corroboraron la especialización funcional de ciertas regiones del neocórtex y constituyeron un gran soporte para el modelo localizacionista. Este modelo ponía de manifiesto las relaciones entre el cerebro, los procesos cognitivos y la conducta humana, y permitía establecer una relación entre algunas áreas lesionadas y determina-

das funciones cognitivas. Este modelo no pretendía quedarse solo en un plano meramente teórico, sino que utilizaba estas relaciones para entender las enfermedades neurológicas y psiquiátricas. De esta manera, se podría pensar que los síntomas de una determinada enfermedad neurodegenerativa se deberían a la atrofia de una u otra estructura cerebral, o incluso que una enfermedad psiquiátrica estaría causada por una degeneración de un lóbulo cerebral específico.

Este modelo de la organización funcional del cerebro sigue siendo hoy aceptado por una parte importante de investigadores y profesionales de la salud. Sin embargo, hay múltiples fenómenos que indican que la especialización funcional es un modelo fallido. Fenómenos como la plasticidad o la diasquisis no pueden ser explicados fácilmente bajo este concepto. Si un área cerebral está diseñada para sustentar un determinado proceso cognitivo, su lesión provocaría una pérdida funcional permanente y no sería posible una recuperación funcional. Por tanto, se negaría la posibilidad de que las intervenciones psicológicas en general, y las neuropsicológicas en particular, pudieran tener sentido. Esto llevó a muchas décadas de nihilismo terapéutico por parte de múltiples profesionales de la salud sobre la eficacia de las terapias psicológicas. Sin embargo, es bien conocido que el cerebro, y por ende el sistema cognitivo, son sistemas plásticos y dinámicos y, por lo tanto, están en constante cambio y adaptación a las exigencias del ambiente. Ante una pérdida funcional, el sistema muestra plasticidad, reorganizando sus propiedades funcionales. Esta reorganización se produce de manera más efectiva cuando se guía mediante un programa de rehabilitación cognitiva. Para interpretar este fenómeno de reorganización, el modelo localizacionista predecía que las regiones periféricas a la lesión o quizás las regiones del hemisferio contralateral serían capaces de asumir la función. Sin embargo, existen múltiples evidencias de que la plasticidad es un fenómeno mucho más complejo que conlleva la reorganización de amplias redes cerebrales que no se limitan a las regiones periféricas o contralaterales.

Los trabajos científicos evidencian que la plasticidad no se produce en regiones específicas, sino que es producto de una reorganización funcional más global en la que se modifican los pesos de las redes y se generan nuevas sinapsis con el objetivo de compensar o sustituir las pérdidas funcionales provocadas por una lesión cerebral. Estos procesos pueden ser lentos, ya que la nueva organización funcional resultante del proceso de plasticidad tiene que llegar a ser funcionalmente efectiva. Así, es lógico que en general la intervención psicológica sea lenta. Debemos pensar que los psicólogos trabajamos en el ámbito de la reorganización anátomo-funcional de las redes funcionales y esto siempre conlleva un tiempo

biológico que suele ser de meses. Así, cuando tratamos una depresión, una fobia o un déficit de memoria son intervenciones lentas. Muchas veces gustamos decir, que la intervención psicológica en general, trabaja reformateando, modificando la red y sus pesos de relación funcional. Sin embargo, la farmacología trabaja más en el nivel de la sinapsis mediante agonistas/antagonistas, produciendo un efecto relativamente más rápido. Mientras que las modificaciones de red ejercidas por la intervención psicológica tienden a producir cambios que se mantendrán en el tiempo, la retirada de un fármaco puede llevar a que la misma red vuelva a la misma dinámica anterior y, por tanto, cause recidivas de los síntomas psíquicos. Así, podemos decir que la teoría de redes sustenta mejor la idea de lo que hacemos los psicólogos, que la teoría localizacionista. Los psicólogos somos terapeutas de redes funcionales y, por tanto, de su propiedad emergente que es la conducta.

La diasquisis es otro fenómeno que es difícil de encajar dentro del modelo de especialización funcional. Es el fenómeno por el que una lesión en un lugar del cerebro produce alteraciones cognitivas que se atribuyen, tradicionalmente, a una región distante del lugar de la lesión. Por ejemplo, lesiones en el cerebelo podrían generar disfunción ejecutiva, un síntoma que aparece típicamente en lesiones prefrontales. Esta idea fue propuesta en 1914 por el neuropatólogo ruso Constantin von Monakow, quien interpretaba este fenómeno como una depresión de la actividad de una región distante, por la pérdida de otra región con la que mantiene una íntima relación funcional. Esta definición implicaba el concepto de que estas dos regiones estarían relacionadas funcionalmente y por tanto formaban una red. Así al alterarse un parte de la red la otra mostraba una depresión funcional.

### **Circuitos o redes funcionales**

Según una corriente, que tiene su origen en los descubrimientos de Santiago Ramón y Cajal a finales del siglo XIX, el cerebro se organiza en circuitos o redes funcionales y, por lo tanto, su alta capacidad computacional no viene dada por las propiedades de sus partes, sino por la relación entre ellas. En el siglo XX, el neurocientífico español Rafael Lorente de N ó, destacado discípulo de Ramón y Cajal y uno de los pioneros de la neurofisiología moderna, describió los primeros circuitos de la vía acústica. Quizá, la mayor contribución de este investigador fue el descubrimiento de los "circuitos cerrados y reverberantes", para los cuales llegó a definir una serie de leyes: una estimulación fuertemente repetida daría lugar al lento desarrollo de una agrupación celular, que a su vez

produciría una facilitación en otros sistemas del mismo tipo. La base de dicho cambio de facilitación de una célula a otra sería la creación de una zona de perturbación local o del aumento de contacto entre las células aferentes activadas y el soma eferente, que produciría el crecimiento de los bulbos sinápticos. Dicho mecanismo sería selectivo y transitorio, basado en una “transmisión opcional” mediante la cual, la dirección en la que se conduciría una excitación dependería de la “sincronización temporal” con otras excitaciones. Los estudios de Rafael Lorente de Nó fueron determinantes para que el psicólogo canadiense Donald Hebb describiera los principios de funcionamiento neuronal en red. Uno de los más conocidos fue el enunciado de que si las neuronas que disparan a la vez tienden a reforzar sus conexiones (*cells that fire together wire together*). Así, y citando literalmente los trabajos de Lorente de Nó, destacó la importancia de la sincronía celular y la organización de la conducta. Esta idea aparentemente simple, daba lugar a lo que se denominó “asambleas neuronales”. Una célula A excita, mediante transmisión sináptica, a una célula B y, por lo tanto, genera una relación de causalidad entre ambas. Cuando este fenómeno se repite en múltiples células se genera una fuerza atractora entre ellas creando una asamblea neuronal. Un aspecto importante de esta teoría sería la idea del aprendizaje asociativo o lo que se denominó “aprendizaje hebbiano”. En este tipo de relación se requiere una activación sincrónica de las células que dará lugar a un incremento de la fuerza de sus conexiones. Así, dos células o grupo de células que repetidamente se activan de forma sincrónica se asocian entre sí y finalmente la actividad en una de ellas facilita la actividad en la otra. Estas ideas coincidían con la existencia de “engramas”, término acuñado a principios del siglo XX por el biólogo alemán Richard Semon para referirse al conjunto de conexiones entre células que contienen información específica en memoria. Los cambios de los pesos en esas conexiones darían lugar a cambios cualitativos de la información soportada. Semon llegaría a decir que “cada estado psicológico corresponde a cambios en los nervios”. El engrama se reactivaría cuando un estímulo exterior recordase un componente del estímulo original que dio pie a su formación. A partir de estas ideas empezaron a formularse hipótesis sobre el funcionamiento de las redes neuronales que posteriormente, con el desarrollo de la física y las matemáticas y su aplicación a la neurociencia, permitieron generar modelos computacionales. Un ejemplo de ello podrían ser los modelos en los que cada peso de conexión se actualiza utilizando, en cualquier lado de la conexión, la información disponible de las neuronas que estén asociadas con ese peso, o para entrenar nuevos patrones tal y como ocurre en las “redes de Hopfield” (que deben su nombre al científico estadounidense John Hopfield), también denominadas “redes neuronales asociativas”.

El modelo del conexionismo, ya explicado más en detalle en un apartado anterior, también intentó trasladar las ideas de los circuitos a modelos computacionales. Estos autores proponían la existencia de procesamiento en paralelo y distribuido a lo largo de múltiples unidades. Para ello, establecieron las reglas de activación de cada unidad, un patrón de conectividad entre las unidades y una ley de propagación de las activaciones de las unidades conectadas entre sí. Estas redes tendrían capacidad de aprendizaje modificando los pesos entre las conexiones. El conexionismo supuso una ruptura conceptual con las ideas previas del funcionamiento del cerebro. La información estaba ahora registrada en amplias capas de conexiones neuronales y en las relaciones de pesos y sus propiedades. Estos modelos computacionales se limitaban a pequeñas poblaciones de neuronas y requerían evidencias experimentales desde los estudios fisiológicos. Por tanto, hacía falta un modelo más global del funcionamiento cerebral, que fuera plausible biológicamente.

Los estudios del fisiólogo estadounidense Norman Geschwind proporcionaron nuevas evidencias del modelo del funcionamiento del cerebro basado en circuitos de procesamiento. En sus estudios pudo demostrar que las lesiones causadas en la sustancia blanca podrían reproducir los síntomas cognitivos encontrados en lesiones de la sustancia gris y que, por tanto, la conducta y los procesos cognitivos se sustentaban sobre las conexiones entre las regiones cerebrales. Geschwind vislumbró la existencia de regiones de alta densidad de conectividad, lo que hoy en día denominamos como *hubs*, cuando atribuía la desconexión entre las regiones corticales a las lesiones en la sustancia blanca o a los daños en las estaciones de obligatorio paso. En la actualidad sabemos que las estaciones de obligatorio paso o *hubs* son esenciales en cualquier modelo del cerebro. No contienen información, pero permiten que las relaciones entre diferentes partes de la red sean fluidas y eficientes. Otros autores, como el neurocientífico español Joaquín Fuster, han destacado como defensores de las teorías de las redes funcionales en sus estudios sobre fisiología animal. Un ejemplo de las relaciones en red sería la memoria, en la que sus tres formas fundamentales (memoria a largo plazo, a corto plazo y memoria operativa) comparten las mismas redes funcionales. De este modo, la memoria operativa serviría como un sistema que permitiría reactivar temporalmente ciertos engramas de la memoria a largo plazo para que se adaptaran a las nuevas informaciones que enriquecen el conocimiento. En esencia Joaquín Fuster afirma que las funciones cognitivas, en general, están sustentadas por las interacciones intra y extra-redes corticales a las que denominará “cognits”.

A pesar de estas evidencias procedentes de la fisiología y de los desarrollos de los modelos computacionales, el modelo localizacionista predominó durante décadas. A finales del siglo XX, el desarrollo de las técnicas de neuroimagen, por un lado, y de las métricas de análisis de la actividad cerebral, por otro, contribuyeron a que se produjera un giro “copernicano” en los modelos que pretenden entender la organización funcional del cerebro. Estas nuevas métricas de análisis matemático de las señales cerebrales permitieron desarrollar ideas como la “conectividad funcional” (termino presentado inicialmente por el neurocientífico británico Karl Friston en 1993; Friston, 2002) o la dependencia estadística entre series temporales. Ejemplos de estas series temporales serían las señales derivadas de la electroencefalografía (EEG) o de la MEG (equipo que será instalado en el campus de Somosaguas en breve), que oscilan a diferentes frecuencias. Filtrando la señal, en diferentes bandas de frecuencia se puede calcular las relaciones de sincronía de fase (por ejemplo, utilizando la transformada de Hilbert) y, por tanto, su dependencia estadística basada en este parámetro. Esa sincronía, estable estadísticamente a lo largo de diversos ensayos de una tarea o tramos de la actividad basal, se interpretaría como una relación funcional entre regiones cerebrales. De esta manera, se pueden calcular enlaces entre nodos o regiones del cerebro a múltiples frecuencias y en diferentes momentos del tiempo, dependiendo siempre de la resolución temporal que disponga la técnica de neuroimagen que se utilice (en el caso de EEG y MEG del orden del milisegundo). Otra posibilidad es que las relaciones entre nodos presenten cierta direccionalidad o “conectividad efectiva”. Es decir, cuando los datos que contiene una serie temporal predicen los de otra serie temporal, podemos decir que una influye en la actividad de la otra o que presentan conectividad efectiva. Este cálculo nos permite tener redes con direccionalidad, lo que las hace más complejas, pero al mismo tiempo más apasionantes.

Así, la conectividad funcional y efectiva han permitido medir, con técnicas de neuroimagen como la resonancia magnética funcional (RMf), la MEG o el EEG, las relaciones entre regiones cerebrales, lo que ha catapultado la investigación sobre las redes funcionales durante la realización de tareas cognitivas. La enorme cantidad de información que ofrecen las redes funcionales en tiempo, frecuencia y número de nodos hacía realmente compleja la posibilidad de asumir, entender y desarrollar un modelo a partir de estos datos. Con la RMf, técnica con una gran resolución espacial, pero con una pobre resolución en tiempo-frecuencia, se ha podido reducir la dimensionalidad de la información y describir una serie de redes que han tenido una enorme relevancia para la comprensión de la organización funcional del cerebro y sus relaciones con la cognición.

Una de las redes que ha cobrado más relevancia ha sido la denominada “red por defecto” (DMN, del inglés *default mode network*). Descrita por el neurólogo estadounidense Marcus Raichle representa la conectividad funcional entre un conjunto de nodos (región medial prefrontal, el cíngulo posterior, el precuneus, las cortezas lateral e inferior parietal y la región medial temporal). Esta red incrementaría su conectividad en estados de ausencia de una tarea cognitiva específica, pero sus valores de conectividad funcional disminuirían cuando el sujeto tiene que hacer una tarea (por ejemplo, restar mentalmente). Una vez que el sujeto deja de hacer la tarea, la red tiende a reconectarse. Se sabe que la desconexión de esta red es esencial para realizar correctamente una tarea cognitiva dada. Las funciones de la red por defecto no se conocen completamente, pero se han asociado con la referencia personal (del inglés *self-reference*), el estado de mente errante (del inglés *mind wondering*), la evaluación social. Pero en los últimos años se ha relacionado más con los periodos de consolidación de la memoria episódica. En algunas patologías neurológicas la red por defecto (DMN) tiende a estar hiperconectada y pierde dinámica (flexibilidad). Es interesante destacar que las regiones típicamente asociadas a la red por defecto son aquellas en las que se produce la acumulación de la proteína beta-amiloide en la enfermedad de Alzheimer, por lo que se ha establecido una asociación entre la DMN y la red de amiloide. Se han descrito otras redes funcionales asociadas a procesos cognitivos, como la red de lo saliente (del inglés *salience network*) o la red ejecutiva. La red de lo saliente se refiere a aquello que tiene la cualidad de sobresalir respecto a otros estímulos. Esto se puede deber a múltiples causas, como el color, la frecuencia y la intensidad de un estímulo (características de bajo nivel del estímulo), la emoción o lo relevante que sea la información desde el plano cognitivo (características de alto nivel del estímulo). El sistema cognitivo se sustenta en este circuito cerebral para evaluar esta saliencia de los estímulos externos o internos dada su relevancia para la adaptación y supervivencia. Así, podemos decir que este sistema es esencial para la identificación de aquellos estímulos biológica o cognitivamente relevantes que permiten guiar flexiblemente nuestra conducta y nuestros recursos de procesamiento. Las áreas cerebrales involucradas en este circuito son: la corteza cíngulada anterior, la región anterior de la ínsula y la región frontal inferior. También involucra algunas regiones no corticales como la amígdala, el estriado ventral y la sustancia nigra / área tegmental ventral. Esta red está implicada en diversas acciones cognitivas como la comunicación, la conducta social, lo emocional y el procesamiento de información cognitiva.

La red ejecutiva dirige la decisión entre otras redes o procesamientos que sean relativamente independientes desde el punto de vista funcional. Facilita la in-

tegración entre la información externa, mediada por la red atencional dorsal, y la información interna, mediada por la red por defecto. Los que aún interpretan el funcionamiento cerebral bajo el paradigma localizacionista esperarían que la red ejecutiva se centrara en las regiones prefrontales. Sin embargo, se sabe que esta red conecta regiones anteroposteriores como la lateral prefrontal, la corteza cingulada anterior y el lóbulo parietal inferior. Por lo tanto, la función ejecutiva no se encuentra localizada en ninguna de estas regiones, sino que está representada en esa red.

Las dos redes citadas anteriormente parecen tener un papel de moduladores y reguladores de la conducta. Hemos podido comprobar que las redes descritas comparten algunos de los nodos, por lo que resulta difícil pensar que estos circuitos sean disociables los unos de los otros. Una de las estructuras, no la única, involucrada en múltiples redes y procesos cognitivos es la corteza cingulada anterior. De hecho, esta región cerebral aparece citada en multitud de publicaciones de “neurociencia cognitiva”. Es una de esas regiones que tienen una alta densidad de conexiones o que se podrían considerar como nodos conectores entre redes, *hubs*.

### **La conectividad anatómica y la conectividad funcional**

Un aspecto que sorprendía inicialmente a los investigadores y que generaba cierta incredulidad acerca de este modelo del conectoma, era la aparente falta de correlación entre la conectividad anatómica y la conectividad funcional. Bajo el paradigma de que las relaciones funcionales en el cerebro deben sustentarse a través de conexiones directas celulares, no se podía interpretar el hecho de que se pueda producir conectividad funcional sin que se dé conectividad anatómica. Estas relaciones funcionales se pueden establecer a través de terceros elementos. De esta manera, podríamos entender que las dinámicas que ocurren en el cerebro son como las que se pueden producir en una orquesta de música. En el cerebro existen directores de orquesta capaces de poner en sincronía diversos grupos celulares a corta o larga distancia, incluso sin que entre ellos exista una conexión directa. De este modo, esas neuronas que no tienen una conexión directa entre sí, pero siguen escuchando a su *hub* conector, como el tálamo, serán capaces de generar pensamiento a través de esa melodía sincrónica y organizada que son las oscilaciones cerebrales. Podemos considerar, desde una perspectiva evolutiva, el origen de las diferencias entre la conectividad anatómica y la funcional. Nuestra adaptación a las exigencias del medioambiente requiere de respuestas adaptativas rápidas. Estas exigencias externas precisan de mecanismos que

generen plasticidad casi inmediata, que podría estar sustentada en un dinámico cambio del peso de las conexiones funcionales y en sus valores de sincronía de fase. Estos cambios, permitirían reclutar y ensayar nuevas conexiones para desarrollar conductas más adaptativas. Estos mecanismos de conectividad funcional inducen mayor vulnerabilidad y probabilidad de interferencia. Por otro lado, la conectividad anatómica sustenta redes que es necesario que sean estables y duraderas, pero también pueden suponer un vestigio del desarrollo de nuestra especie que incluso posean un cierto determinismo genético. Algunos autores han llegado a postular que aquellas habilidades cognitivas, que sean de uso frecuente y exitoso por nuestro grado de especialización, podrían comenzar a desarrollarse como conexiones funcionales, creadas por un sistema de ensayo-error, para después generar conexiones anatómicas que las harían subsistir disminuyendo su vulnerabilidad. De esta manera, estaríamos describiendo fenómenos de plasticidad rápida (funcionales) y otros más lentos y orientados a largo plazo mediante una conectividad anatómica. En cualquier caso, resulta indudable que la conectividad estructural no es suficiente para justificar la complejidad del pensamiento humano. Estas relaciones anatómico-funcionales se antojan más complejas de lo que hemos expresado anteriormente y serán clarificadas por las investigaciones del Proyecto Conectoma Humano. Este proyecto de investigación multicéntrico, que actualmente involucra a investigadores internacionales, se ha puesto como meta describir el conectoma anatómico y funcional, e indicar su alteración en algunas patologías neurológicas como la enfermedad de Alzheimer. La complejidad de las conexiones anatómicas y funcionales desorientaba a los investigadores que proponían un sinnúmero de interpretaciones que difícilmente llegaban a un modelo basado en unos principios sólidos y demostrables empíricamente. Se hacía necesario desarrollar una teoría que pudiera ayudar a entender el carácter “sociológico” de las relaciones entre neuronas y regiones cerebrales, con una visión de macroescala en la que se perdieran las referencias anatómicas clásicas y se buscaran principios de comportamiento de las arquitecturas funcionales. Así, muchos investigadores se han visto atraídos por el modelo matemático de la “teoría de grafos” que proporcionaría una comprensión más holística del funcionamiento cerebral.

### **La teoría de grafos y su aplicación en psicología y neurociencia cognitiva**

Imaginemos que cada persona del mundo es una neurona en un cerebro que podríamos representar como el planeta Tierra. Las personas tienden a estar conectadas (relacionadas con mayor frecuencia) con aquellas personas de

las que les separan una menor distancia física, a las que ven todos los días. Pero algunas viajan mucho, de unos países a otros, y conocen a individuos de diferentes lugares del mundo. Son capaces de transmitir información de unas comunidades a otras y de esa manera romper la endogamia en la información que se genera en una comunidad, haciendo que los descubrimientos e ideas sean fácilmente transferibles de unos lugares a otros. Esta función de transferencia de información ha permitido que la humanidad avanzara más rápidamente, ya que el conocimiento fluye gracias a esos nodos o individuos conectores entre culturas. Este tipo de relaciones no solo se establecen en el plano puramente físico, sino que también las podríamos modelizar en las relaciones establecidas en una red social de internet. Si los miembros de un grupo particular dentro de una clásica red social solo se enviaran mensajes entre sí, la información sería muy redundante (en los temas y en las imágenes). Sin embargo, si alguno de los miembros de ese grupo perteneciera a muchos otros grupos dentro de una misma red social o en diversas redes, su capacidad de eliminar la redundancia sería crucial ya que podría introducir nueva información y dinámicas variadas en una comunidad particular. La Psicología Social ya tuvo esta idea en la década de 1960 y empezó a utilizar estos criterios para entender las relaciones sociales.

Todas estas ideas se pueden modelizar dentro de una teoría matemática denominada "teoría de grafos". En el siglo XVIII, el matemático suizo Leonard Euler vivía en una ciudad de la antigua Prusia denominada Königsberg (conocida hoy como Kaliningrado en Rusia), donde había siete puentes que cruzaban el río Pregel y unían las áreas de la ciudad de cada orilla con dos islas. Todo el mundo, se preguntaba si se podría pasear por la ciudad cruzando solo una vez cada uno de los puentes. Euler, en vez de recorrer infinitas veces la ciudad para resolver el problema, dibujó un grafo donde cada rivera y cada isla fueran un nodo y los puentes, sus enlaces. Así representó cuatro nodos (dos riveras y dos islas) y siete enlaces, llegando a la conclusión de que más de dos nodos de la red no deberían tener un número de enlaces impar, que les conectara con el resto de los nodos del grafo, para poder cruzar toda una sola vez. De esa manera, demostró que no se podía pasar por los puentes de Königsberg una sola vez, ya que todos los nodos tienen un número impar de enlaces. Esta forma de representar en un grafo las relaciones entre nodos de una red se fijaba esencialmente en el número de enlaces y descartaba otras informaciones como la distancia física entre las áreas de la ciudad y, por lo tanto, el modelo podría adecuarse a cualquier transformación física de un espacio cualquiera. Esta idea de generar grafos a partir de nodos y enlaces permitió producir modelos de redes de un alto número de elementos, como es el caso del cerebro. Podríamos

poner como ejemplos el mapa de carreteras de un país, las relaciones en una red social o económica, los ecosistemas, etc. Todos podrían entrar en la definición de lo que es un sistema complejo, formado por un sistema de interacciones entre elementos, cuyas propiedades globales no puedan ser explicadas por cada una de sus partes individualmente. Una de las propiedades de un sistema complejo se puede denominar como emergentes, nuevas, en el sentido de que no se justifican por la suma de sus partes. Los sistemas complejos eficientes no son, ni totalmente aleatorios (todos los nodos tienen la misma probabilidad de conectarse con otro nodo cualquiera de la red) ni totalmente regulares (las conexiones tienden a hacerse con los vecinos más próximos). En un sistema complejo las interacciones añaden una información adicional o propiedad emergente. Así, la estimulación permite generar espontáneamente propiedades emergentes, por la interacción de los elementos y la autoorganización. Sería el caso de un huracán en el que las moléculas de aire en movimiento se ordenan formando su estructura característica. El pensamiento podría considerarse como una propiedad emergente de un sistema complejo autoorganizado, como es el cerebro. También lo serían la creatividad y la solución de problemas, que tienen un claro fin adaptativo. Siguiendo el ejemplo que considera el cerebro como una red social, podríamos decir que está formado por nodos (neuronas o regiones cerebrales) y enlaces (anatómicos o funcionales) que pueden presentar arquitecturas funcionales características. En 1998 el sociólogo canadiense Duncan Watts, de la Universidad de Columbia, y el matemático estadounidense Steven Strogatz, de la Universidad de Cornell, publicaron un famoso artículo en la revista *Nature* en el cual introducían el modelo de "mundo pequeño". El mundo pequeño puede ilustrarse con el ejemplo de las relaciones sociales en las que la distancia aproximada entre dos personas cualesquiera del planeta podría ser de seis (yo conozco a una persona, que conoce a otra y así sucesivamente hasta alcanzar a cualquier otra persona del planeta). Estas arquitecturas de mundo pequeño se caracterizan por establecer un equilibrio entre segregación e integración, de manera que presentan un alto grado de conexiones entre nodos vecinos (agrupación, del inglés *cluster*) además de una baja distancia (número de pasos para llegar de un nodo a otro dentro de la red) entre nodos (por ejemplo, áreas del cerebro). Cuando se aplican las ideas de Watts y Strogatz a los análisis de datos en neurociencias, de alguna manera se supera la concepción tradicional de las áreas cerebrales, ya que en un grafo las distancias físicas se minimizan para pasar a evaluarse las distancias por el número de nodos que tenemos que atravesar para llegar de una parte a otra de la red. La organización de mundo pequeño estaría contrapuesta a una red de arquitectura aleatoria (todo conectado con todo) o regular (sólo me conecto con los vecinos

más próximos). Además, las redes de mundo pequeño son más eficientes en el consumo energético que las redes aleatorias y también más eficientes en el procesamiento de información que las redes regulares, estableciendo un buen equilibrio entre eficiencia y gasto energético. Si observáramos al cerebro desde la perspectiva de la economía, sería una empresa sumamente eficiente ya que presenta una alta productividad minimizando el gasto.

### **La importancia de los *hubs* en las redes cerebrales**

La teoría de grafos añadió un elemento esencial para la comprensión de la estructura de las redes funcionales cerebrales: la descripción de los *hubs* o áreas de alta densidad de conectividad. Estas regiones son esenciales para el ahorro energético, tanto en las redes estructurales como en las redes funcionales. Además, sirven de enlace entre múltiples áreas de una misma comunidad o de nodos pertenecientes a diferentes comunidades. El daño causado en un *hub* podría provocar alteraciones en la conectividad de una red e importantes trastornos cognitivos, mientras que el daño en un nodo periférico que no fuera un *hub* generaría un menor deterioro cognitivo. Quizá sea este el punto de encuentro entre la teoría localizacionista y el modelo de redes. El hecho de que se genere un síndrome neuropsicológico al dañarse un *hub* llevó a los neurocientíficos a interpretar que ahí se localizaba una función cognitiva particular. Sin embargo, la pérdida de un nodo que permite que la información fluya a lo largo de la red, producirá una importante alteración de toda la red y, por lo tanto, de la función cognitiva que sustenta. La probabilidad de encontrar nodos con alta densidad de conectividad es más elevada en las redes de escala libre, lo que indica una fuerte organización de la arquitectura funcional que se aleja de una estructura aleatoria. De este modo, cualquier modelo computacional (o inteligencia artificial) del funcionamiento cerebral necesitaría implementar la existencia de *hubs* para poder considerarse biológicamente plausible. Esto no es así completamente en los modelos actuales de IA. Cuando se establece un histograma de distribución de las conexiones de cada nodo, se puede observar que el número de nodos que se pueden considerar *hubs* es siempre mucho menor que los que presentan menos conexiones o *no-hubs*. Todas las redes descritas hasta ahora como la DMN presentan nodos de alta densidad de conectividad y los modelos de emulación del cerebro están evaluando la importancia de estos nodos en las redes funcionales y estructurales. Su implementación en la IA podría generar un importante salto en la velocidad de procesamiento de los cálculos de las redes artificiales. Este será un nuevo campo de inspiración desde la psicología cognitiva a la IA.

## **Entender las enfermedades neurodegenerativas a partir de los modelos del conectoma**

Una vez que se acepta que el cerebro funciona con una organización de redes funcionales que pueden ser calculadas mediante las métricas de conectividad funcional y que pueden ser modelizadas mediante la teoría de grafos, se plantea de inmediato la pregunta de si esta aproximación podría ser aplicada al diagnóstico de enfermedades cerebrales e incluso a su tratamiento. El diagnóstico de una enfermedad neurológica o psiquiátrica requiere de la asociación de un conjunto de signos y síntomas, donde los patrones de las redes cerebrales serán una pieza más del algoritmo de diagnóstico. Por lo tanto, en sí mismas podrían ofrecer cierta sensibilidad y especificidad, pero el diagnóstico clínico es algo más complejo. Ciertamente, en algunas enfermedades o bien no se obtiene un diagnóstico certero o se alcanza relativamente tarde. En estos síndromes parece necesaria la inclusión de una nueva perspectiva que aporte nuevos datos. La aproximación de las redes funcionales puede interpretar de otra forma gran parte de los fenómenos que quedaban sin justificar mediante el procedimiento localizacionista y que afectan a la comprensión de los síndromes neuropsiquiátricos. Así, cuando un sistema se daña se produce: degeneración transneuronal, diasquisis y desdiferenciación. Además de estos fenómenos que representan el daño funcional, existen otros fenómenos positivos como: la compensación, la reserva cognitiva o la sustitución. La degeneración transneuronal se refiere a la habilidad de la red cerebral para transmitir ciertas características patológicas a lo largo de una red funcional. Este fenómeno se está describiendo como una novedosa posibilidad en la enfermedad de Alzheimer, donde las neuronas se podrían transmitir unas a otras la neuropatología de la enfermedad utilizando las redes cerebrales. Ya describimos anteriormente el fenómeno de la diasquisis, que muestra cómo la alteración en una parte de la red generaría cambios de dinámica en la globalidad del circuito. Por último, la pérdida de la especificidad de las redes o de su especialización funcional podría alterar el carácter modular y las arquitecturas de tipo pequeño mundo (desdiferenciación). Para tratar de resolver estos problemas que pueden darse en diferentes enfermedades neuropsiquiátricas se pueden dar una serie de fenómenos como la compensación (incremento de la actividad de la red en regiones no dañadas). La plasticidad, cambiando los pesos sinápticos, generando nuevas vías o destruyendo conexiones espurias, permitiría compensar la pérdida de nodos o enlaces en la red. La reserva cognitiva se basaría en la disponibilidad de una robusta arquitectura anatómico-funcional que permitiría resistir a la progresiva destrucción de enlaces o nodos, por ejemplo, en una enfermedad neurodegenerativa. Estas redes se establecerían mediante entrenamiento cognitivo, y también se beneficiarían del ejercicio físico y de una nutrición (por ejemplo, la dieta

mediterránea). El último caso sería el de que otros sistemas asumirían la función dañada o sustitución. Sería un elemento típico en el neurodesarrollo cuando la plasticidad del sistema permite máxima flexibilidad, generando alternativas más allá de las tendencias organizativas de la red. Existen múltiples patologías cerebrales donde ya se han caracterizado las alteraciones específicas de red. Quizá gran parte de estos desórdenes se podrían agrupar en dos grandes categorías, que no son excluyentes: los síndromes de desconexión y la pérdida de balance entre excitación e inhibición. Los síndromes de desconexión serán aquellos en los que predomine un daño en los enlaces entre nodos o en los *hubs* de una red. Este podría ser el caso de la esclerosis múltiple, donde los daños en la sustancia blanca impedirían una conexión fluida entre nodos, así como de algunos síndromes del neurodesarrollo, donde el estudio mediante tractografía por resonancia magnética está permitiendo observar alteraciones congénitas o adquiridas en el normal desarrollo de los circuitos cerebrales. Otro factor que altera el normal funcionamiento de las redes funcionales es la pérdida del balance entre excitación e inhibición. Este fenómeno puede tener lugar en aquellas patologías donde el daño selectivo en las neuronas inhibitorias produzca un excesivo disparo de las células excitatorias y genere excesiva sincronía funcional sin que tenga un papel adaptativo. Este es el caso de la epilepsia, donde el exceso de sincronía reduce enormemente la entropía, o de la esquizofrenia, donde las conexiones gabaérgicas han perdido protagonismo y se desestabiliza la red funcional. La enfermedad de Alzheimer ha sido tradicionalmente estudiada desde una perspectiva puramente localizacionista, en la que se buscaba la atrofia de regiones mediales temporales o áreas de acumulación de proteínas (amiloide y tau). Sin embargo, desde la teoría del conectoma se está tomando una perspectiva algo diferente. La alteración de los procesos cognitivos y los cambios conductuales se darían como efecto de la progresiva pérdida de espinas dendríticas y la alteración del balance entre excitación e inhibición. Los procesos neuropatológicos de la fosforilización de la proteína tau y la acumulación de la proteína beta-amiloide dañan la conectividad y por tanto se podrían caracterizar como un síndrome de desconexión. Sin embargo, en la última década se está observando, tanto en registros de resonancia magnética funcional como de EEG y MEG, un incremento en los valores de la conectividad funcional. En el caso de los registros neurofisiológicos se caracteriza como una hipersincronización y, por lo tanto, un incremento de los valores de los enlaces de la red. Este excesivo incremento de la sincronía provoca la pérdida de la estructura normal de red de escala libre y una alteración de los *hubs*, haciendo que la red tienda hacia una estructura aleatoria. Esta excesiva conectividad facilitaría además la transmisión, neurona a neurona, de la fisiopatología de la enfermedad generando una difusión rápida por todo el cerebro de las formas

oligoméricas de tau. La compresión de muchos de estos síndromes, como una pérdida del balance excitación-inhibición, está llevando a intentar otras posibilidades farmacológicas y no farmacológicas que restauren estos desequilibrios. Basándonos en esta hipótesis nuestro grupo realizó registros MEG antes y después de una intervención neuropsicológica observando los siguientes fenómenos: 1) el grupo sin intervención cognitiva tendía a incrementar la excitabilidad de la red en el tiempo; 2) el grupo que recibió la intervención cognitiva estabilizó ese balance y mejoró en su rendimiento neuropsicológico; 3) la intervención cognitiva recuperaba *hubs* cerebrales que se estaban dañando. Así, podemos decir una vez más que la intervención psicológica induce importantes cambios funcionales en la red cerebral, en la homeostasis excitación/inhibición y más importante en el funcionamiento cognitivo de los pacientes con deterioro cognitivo.

### **Tratamiento de los desórdenes neurológicos con la aproximación del conectoma y la ayuda de la IA y la Neurociencia Computacional**

Una vez que se conoce la red de una determinada patología, se plantean algunas posibilidades en las que los modelos de IA pueden servir de ayuda: la clasificación, la predicción y la intervención. La clasificación permite extraer conjuntos de características que permitan establecer patrones diferenciables entre personas con diferentes patologías o clasificaciones de imágenes o de patrones de ejecución de pruebas cognitivas. La predicción se basaría en la posibilidad de describir futuros estados de la red y, por lo tanto, las alteraciones que esta sufriría en el tiempo. Esto se podría hacer utilizando el aprendizaje automático (*machine learning*), todavía poco empleado para esta finalidad, o modelos mecanísticos (de explicación de mecanismos) desarrollados en el contexto de la neurociencia computacional. Los modelos de “*spiking*” o los modelos de masas neuronales (*neural mass models*) han sido los más utilizados para modelar la actividad cerebral. La idea es que estas métricas sean capaces de simular matemáticamente los parámetros que rigen una determinada estructura de red para después poder “testar” in-silico posibles intervenciones. Así, estos modelos computacionales de las redes de las enfermedades neuropsiquiátricas servirán como plataformas para realizar cambios en la estructura, la dinámica y las propiedades de conectividad de dichas redes, con el objetivo de revertir los estados patológicos y alcanzar arquitecturas funcionales cercanas a la normalidad.

Un ejemplo práctico que se está implementando en muchas clínicas de psicología es la combinación de la intervención psicológica con modelos de neuromodulación por corriente de bajo voltaje. Para saber el lugar, la frecuencia e intensidad

de estimulación son esenciales los modelos de neurociencia computacional. Estos modelos están siendo desarrollados actualmente en la Facultad de Psicología de la UCM, gracias a la disponibilidad de datos reales de MEG. Esta metodología es aplicable a trastornos psíquicos y cognitivos y será una magnífica línea de trabajo entre la clínica de psicología, la clínica de logopedia, el Centro de Neurociencia Cognitiva y Computacional, todos de la UCM, y el laboratorio de Psicología Experimental.

## **Desde la Psicología Cognitiva a la Inteligencia Artificial**

Los psicólogos cognitivos han desarrollado modelos de los procesos cognitivos mediante experimentación empírica en sujetos normales, además de apoyarse en la aplicación de tareas cognitivas específicas en pacientes con lesiones cerebrales. Adicionalmente, los estudios con técnicas de neuroimagen y de neuromodulación han aportado toda una nueva fuente de datos que ha permitido el establecimiento de un conocimiento sobre el sistema cognitivo que va desde lo fisiológico a lo cognitivo-sintáctico. Así se han formalizado reglas y predicciones formando constructos y modelos que nos permiten entender mejor la mente humana. Incluso se ha llegado a generar una disciplina que es la Psicología del Pensamiento, asignatura que impartimos en nuestro departamento, en la que se enseñan los modelos sobre el sistema semántico, sobre el razonamiento (deductivo e inductivo) y la resolución de problemas. Este conocimiento tan profundo de la mente humana es la base sobre la que se ha desarrollado en gran medida la ciencia computacional.

Quizá uno de los conceptos que más ha influido en los modelos bio-inspirados de IA, es la idea de la psicología cognitiva de que el cerebro humano es un “procesador de información”. A partir de la codificación de los datos del entorno, con ciertos sistemas de selección de información, se realizarán una serie de computaciones dando lugar a unas salidas o respuestas. Las sucesivas computaciones generaran modificaciones en el sistema cognitivo, de manera que provocaran un aprendizaje continuo (al menos en la fase de entrenamiento).

Una vez que los primeros modelos empíricos de la cognición fueron formalizados, se planteó la analogía cerebro-ordenador. Así la mente es al cerebro lo que

el software sería al hardware del ordenador. De esta manera, cualquier sistema físico que tuviera el programa correcto podría tener una mente igual a la de un ser humano. El problema era construir un software capaz de imitar la complejidad de las computaciones cognitivas y el sistema de representación de información. Quizá hubiera un problema más difícil de superar, si la mente es una propiedad emergente del cerebro, ¿cómo podríamos generar propiedades emergentes con un ordenador? Sin embargo, la investigación en psicología cognitiva siguió avanzando y se realizaron descubrimientos esenciales que inspiraron y siguen inspirando muchos de los modelos de IA. En el capítulo sobre memoria, avanzamos ya las ideas conexionistas. Los psicólogos cognitivos conexionistas revolucionaron la manera de entender la organización de la mente humana, incluso se atrevieron a desarrollar los primeros programas informáticos que la imitaban. Se basaron en las ideas de la organización funcional de la neurona y de las redes de conexiones. La clave está en explicar cómo se realiza el procesamiento de la información del que hablábamos anteriormente. Las entradas de información generarán un patrón propagado de activaciones neuronales (redes neuronales) que van a generar unas salidas específicas. Estas investigaciones realizadas en gran medida por psicólogos, en colaboración con ingenieros informáticos, que generaron diferentes modelos con reglas sintácticas, en su versión más primitiva preprogramadas, pero que progresivamente fueron siendo más automáticas.

Veamos entonces a continuación el paralelismo con los desarrollos de la IA. El inventor del término "Inteligencia Artificial" John McCarthy (acuñó el término en 1956, cuando organiza la mítica conferencia de Dartmouth), dijo "puede decirse que máquinas tan simples como los termostatos tienen creencias". John Searle (1990), filósofo de la ciencia, le preguntó a McCarthy cuáles eran esas creencias y este le contestó: "Mi termostato tiene tres creencias, hace demasiado calor aquí, hace demasiado frío aquí y aquí hace la temperatura correcta". Sin duda, estas "creencias" están determinadas por la información previa introducida en su software sobre que valores significan frío/calor o cual es la temperatura correcta. Así, es un sistema puramente sintáctico que sigue reglas muy estrictas impuestas *a priori*. Sin embargo, McCarthy llegó a decir en esos años "La velocidad y capacidad de memoria de los computadores actuales puede ser insuficiente para simular muchas de las funciones más complejas del cerebro humano, pero el principal obstáculo no es la falta de capacidad de las máquinas, sino nuestra incapacidad de escribir programas que aprovechen por completo lo que tenemos". Desde el comienzo se planteó la idea de que una verdadera IA debía pasar el test de Turing, según el cual un humano realiza preguntas a través de la pantalla de un ordenador, y si no puede decidir si el que le está respondiendo

es otro humano o una máquina, entonces ésta es definitivamente inteligente. En 2014 un *bot* computacional llamado “Eugene Goostman” fue capaz de engañar a 30 de los 150 jueces a los que se sometió durante el test de Turing haciéndoles creer que estaban hablando con un niño ucraniano de 13 años.

Uno de los primeros programas que imitaban la mente humana lo crearon Newell, Sha y Simon en 1957 y lo bautizaron como “General Problem Solver” (GPS; solucionador general de problemas). La idea era utilizar las teorías cognitivas sobre la resolución de problemas y crear un sistema para resolver problemas suficientemente formalizados. Si bien el GPS resolvió problemas bien formalizados, como la Torre de Hanoi, no pudo resolver ningún problema del mundo real, era computacionalmente insostenible. De cualquier forma, el GPS fue capaz de separar su conocimiento de los problemas, de su estrategia de cómo resolverlos.

Posiblemente fue Marvin Lee Minsky (inspirado por McCulloch y Pitts), el que inspirándose en las teorías conexionistas y de redes neuronales provocó uno de los avances más trascendentales en la IA con el desarrollo de SNARC en 1951, que fue el primer simulador de redes neuronales (implementación de una red de 40 neuronas que modelaba el comportamiento de una rata para salir de un laberinto). Su tesis doctoral de 1954 “Redes neuronales y el problema del modelo de cerebro” era ya todo un alegato de la investigación en redes neuronales artificiales. Su investigación inspiró al psicólogo norteamericano Frank Rosenblatt para que implementara el “Perceptrón ” en 1957. Este programa se inspira en la estructura de una neurona con múltiples entradas o dendritas y una salida o axón y en el hecho de que las neuronas forman poblaciones densamente conectadas entre sí. En el caso del Perceptrón, se basaba en una arquitectura de una sola capa, teniendo varias neuronas de entrada y una o varias neuronas de salida, donde cada neurona de entrada se conecta a cada neurona de salida. El objetivo era el de crear un método de clasificación, mediante la determinación de una recta (en el caso de tener dos clases) que separe los ejemplos mostrados a la red en dos regiones del espacio. El proceso de aprendizaje de Perceptrón consiste en un proceso iterativo en el que se van modificando los pesos entre las conexiones que forman estas neuronas, hasta encontrar los valores que forman la mejor recta posible. El valor de las entradas ( $x_i$ , que representaría a la sinapsis biológica) se multiplica por el peso de las conexiones ( $w_i$ ), y posteriormente se realiza un sumatorio, de todos los pares  $x_i * w_i$  (esto es lo que haría la neurona biológica, integrando la información que proviene de todas las sinapsis). Para producir la salida final de la neurona se aplica una función de activación, en el caso de Perceptrón se aplica una función escalón, que hace que la salida de la red sea binaria, lo cual limitaba mucho los

datos a codificar. Por eso Widrow y Hoff desarrollaron ADALINE en 1960, que era una estructura igual a Perceptrón, pero que era capaz de calcular la diferencia real entre entradas y salidas. De esta forma, la red podía saber cuánto se había equivocado en la predicción, al contrario que Perceptrón que solo podía saber si se había equivocado o no. Ni Perceptrón ni Adaline contaban con capas ocultas porque aún no se habían propuesto los métodos matemáticos para entrenar esas capas. Así que ambos casos son ejemplos clásicos muy sencillos.

El contacto entre neuronas se establece normalmente en capas de poblaciones neuronales que reciben y envían información a la siguiente capa. Así, se estableció el concepto de redes neuronales artificiales, en donde una serie de neuronas artificiales (unidad básica de computación) estarían conectadas en capas, de forma que la información de entrada atraviesa la red neuronal, provocando diversas computaciones, y finalmente generando unos valores de salida. En los enlaces entre neuronas el valor de salida de la neurona anterior es multiplicado de nuevo por un valor (peso). Los pesos sirven para “activar” o “inhibir” a las neuronas con las que se conectan. Los valores de salida de cada neurona estarán sometidos de nuevo a umbrales para establecer unos valores máximos. Estas redes de capas tienen la capacidad de aprender automáticamente, intentando minimizar la función de coste, que es un sistema de evaluar la eficacia de la red en su conjunto. Para aprender, se tienen que actualizar los valores de los pesos de los enlaces para ir minimizando la función de pérdida, utilizando la retropropagación (la información de la función de pérdida fluye hacia atrás para recalcular los pesos de los enlaces). La retropropagación se ha criticado desde la neurociencia ya que es un modelo que no parece biológicamente plausible. Fue alrededor de los años 90 cuando se genera el concepto de aprendizaje profundo (*deep learning*). Se crean redes formadas por varias subredes, en las que cada una de ellas presenta un proceso de aprendizaje intrínseco.

Estas ideas han evolucionado hasta desarrollar sistemas expertos de detección y clasificación en temáticas específicas, modelos de lenguaje natural y en los recientes sistemas de Chatbots. Quizá el que se ha hecho más popular, alcanzó más de 100 millones de usuarios en dos meses, es el denominado ChatGPT de la empresa OpenAI, lanzado en su versión más avanzada en noviembre del 2022. El significado de este acrónimo revela mucho de lo que es y hace “Chat Generative Pre-trained Transformer”. La primera palabra se refiere a que es un sistema de conversación “inteligente” (modelos de lenguaje natural). El preentrenamiento generativo se refiere a que este sistema ha sido entrenado con grandes bases de datos (570GB de datos de texto) de libros, artículos, páginas web, redes

sociales. Gracias a este entrenamiento puede predecir cual será la siguiente palabra, tras una secuencia, con gran eficacia. La palabra “*transformer*” se refiere al tipo de arquitectura de redes neuronales sobre el que está basado. En general es dependiente de la información utilizada durante el periodo de entrenamiento, pero podría mejorar alguna de sus limitaciones cuando se le dé acceso al conocimiento general de internet. Este Chat se está usando ya para diversas actividades (programación, búsqueda de información) y es muy probable que muchos de nuestros estudiantes lo hayan hecho para la realización de trabajos de clase. Este que les habla le pidió que realizara un trabajo sobre un tema muy específico de su asignatura y dio como resultado un magnífico resumen sobre el tema en cuestión. Tras pasarlo por nuestro programa de evaluación de posible copia (Turnitin) se daba un resultado de originalidad casi del 90%. Es posible que debamos replantear los objetivos de los trabajos que pedimos a los estudiantes.

## **Sistemas sintácticos y sistemas semánticos, de máquinas y de humanos**

Una vez que hemos revisado cómo la psicología cognitiva ha ido formalizando los modelos de la mente, las nuevas avenidas de la teoría del conectoma y cómo todo esto está inspirando y seguirá inspirando el desarrollo de sistemas artificiales, es importante hacer una reflexión sobre cuáles serían las verdaderas similitudes y diferencias entre la cognición humana e inteligencia artificial.

En 1990 el filósofo de la ciencia John Searle decía: “Tener una mente es algo más que tener procesos formales o sintácticos. (...) La mente tiene más que una sintaxis, tiene una semántica”.

### **Emociones y sistema semántico**

Quizá un elemento esencial para entender al ser humano es la influencia que ejercen las emociones en el sistema cognitivo. Como decíamos en el apartado que le dedicamos a las emociones, estas son una de las mayores fuentes de semántica para nuestro sistema cognitivo, determinando el aprendizaje y ensalzando el procesamiento profundo de la información, tanto implícitamente como declarativamente. La semántica de la emoción influye en algo tan difícil de entender como es el sarcasmo o los chistes (existentes en todas las culturas).

Esta es una de las primeras habilidades semánticas que pierde un enfermo de Alzheimer en su progresivo proceso de deterioro, lo que nos indica su altura en la jerarquía de la complejidad cognitiva.

Las emociones son un elemento muy difícil de formalizar en un sistema de IA. Las aproximaciones de lenguaje natural en IA pueden conseguir que una máquina simule muy bien una conversación emocional de un humano, hay cientos de miles de conversaciones emocionales en Twitter. Sin embargo, si la fuente de información que alimenta el entrenamiento de la IA son las redes sociales, lo más seguro es que el Chatbot sea bastante gruñón, ya que en estas redes sociales predominan los comentarios negativos frente a los positivos. Sin embargo, los seres humanos no somos en líneas generales tan gruñones. Además, nuestras emociones cumplen un patrón estocástico y, por tanto, supone un problema importante para los algoritmos de lenguaje natural hacer predicciones precisas de cuáles serían las palabras adecuadas dentro de la cadena de una frase de contenido emocional. Si alguno de los lectores de este documento ha tenido la fortuna de convivir con un adolescente, podrá entender la idea de lo estocástico que puede llegar a ser un ser humano, son como la caja de bombones de Forrest Gump, nunca sabes lo que te va a tocar. Esta falta de determinismo y el fuerte componente aleatorio de las emociones hace difícil tomar predicciones de conductas en contextos específicos, algo que ha sido un verdadero caballo de batalla en la investigación en psicología.

### **La comunicación no verbal**

Otra de las facetas que puede resultar enormemente compleja de simular por una IA es la comunicación no-verbal. Esta supone más del 80% (el 38% el tono de voz y el 55% el lenguaje corporal; sólo el 7% lo verbal) de nuestra comunicación habitual (Mehrabian, 1971). Muchos juzgamos a las personas por su capacidad de comunicación, no tanto por lo que dice sino cómo lo dice. Es posiblemente lo que diferenciará a un buen profesor de otros. Todos tenemos conocimientos muy parecidos pero la manera de comunicarlos marca una gran diferencia. Así la comunicación no verbal será la trasmisión de información sin hacer el uso del habla, a través de las expresiones faciales, gestos y posturas corporales, la distancia física, el contacto físico, la fijación de la mirada, etc. Es un sistema de comunicación utilizado por todas las especies y en los humanos la información que trasmite suele ser más sincera e incontrolada que los mensajes verbales. Se sabe, por las investigaciones del psicólogo Paul Ekman, que existen hasta 15 expresiones de nuestro rostro que son semejantes en todas las culturas y

que transmiten un mensaje semejante. Lo mismo ocurre con algunas posturas corporales, que incluso compartimos con primates no humanos. En el futuro se tenderá al desarrollo de humanoides con capacidades de comunicación no verbal. En la neuro-robótica cobrarán nuevamente un papel esencial los estudios de las emociones en psicología.

### **La representación semántica: las redes semánticas**

En general formalizar el sistema semántico ha sido tarea difícil para los psicólogos debido a su inmensa complejidad y lo intrincadas que están las diferentes representaciones semánticas las unas con las otras. Aun así, se han llegado a soluciones que han logrado formalizar de manera bastante certera las representaciones del sistema semántico. Un ejemplo de ello es lo que se denominó “redes semánticas”, que han inspirado enormemente los modelos de lenguaje natural y de ontologías en la IA.

Quizá el modelo más conocido de red semántica es el “Teachable Language Comprehender” (TLC), desarrollado por el psicólogo norteamericano Ross Quilliam. Este autor plantea que cada palabra, y su concepto, pueden representarse como un nodo dentro de una red. Cada nodo tendrá asociada una serie de propiedades, en su ejemplo del pájaro “canario” las propiedades serían: puede cantar y es amarillo. Cada nodo, tendrá una relación de preferencia con otros nodos de la red semántica. En el caso del “canario” su relación preferente será con la palabra “pájaro”, que a su vez tendrá las propiedades de: puede volar, tienes alas, tiene pico. Este nodo, “pájaro”, se conectará con “animal” y así sucesivamente se formaban redes semánticas complejas de nodos que forman categorías semánticas. La velocidad de procesamiento dentro de la red de la misma categoría será muy rápida y por tanto responder si la siguiente frase es correcta “El canario puede volar” se realizará con un tiempo de reacción corto. Esto se debe a la “ley de propagación” en la que una vez que se activa un nodo de la red, la información se propaga automáticamente a lo largo de la red y por tanto el procesamiento de todas las propiedades está accesible muy rápidamente.

En la jerarquía de la red semántica, los nodos de más alto nivel serán más generalistas en sus propiedades, mientras los de más bajo nivel serán más específicos. Un ejemplo de lo primero sería la palabra “animal” y de lo segundo “canario”. En los desarrollos posteriores de este modelo se empezaron a poner pesos a los enlaces de la red para poder asumir el hecho de la existencia de nodos prototípicos de la categoría o nodos a los que los participantes respondían sistemáticamente

de forma mucho más rápida que a otros nodos del mismo nivel jerárquico. Los pesos también se adaptaron al efecto de la familiaridad, nodos más familiares se respondían antes, etc.

### **La inferencia semántica**

La representación semántica no aborda todas las propiedades del sistema semántico. Claramente los seres humanos tenemos la capacidad de tener la habilidad de generar información completamente nueva, con una referencia limitada a la información previamente acumulada en el sistema cognitivo. A esta habilidad la llamaremos aquí “inferencia semántica”. Sin duda, la generación de una nueva idea, un innovador concepto requiere de una acumulación de conocimiento que lleve a un punto crítico en el que se produzca un cambio de estado. Por tanto, es un cambio de estado basado en una inferencia participada de la experiencia previa. Es por ello, que en ciencia sea común que diversos científicos en diferentes partes del mundo generen soluciones muy semejantes ante un mismo problema. La globalización del conocimiento, a la que nos ha llevado la aparición de la “gran red” ha incrementado la frecuencia de este fenómeno de coincidencia temporal de genialidades. Picasso creó el cubismo en parte gracias a la información previa del impresionismo y otros estilos, pero el salto entre los dos estilos es claramente exponencial y no es simplemente una relación lineal. Este tipo de funciones exponenciales de la cognición son, hoy en día, imposibles de formalizar y menos de introducir en una IA. Puedo formular el paso de agua a vapor y describir su función matemática. Es un fenómeno que siempre pasa de la misma manera. Sin embargo, la inferencia semántica, no se produce bajo reglas deterministas. La creatividad, en todas sus facetas, sería el ejemplo más claro. La ciencia psicológica e incluso la filosofía lleva centenares de años teorizando sobre la creatividad, pero ni los propios artistas pueden predecir cuándo llegará la genialidad. Picasso decía: “que la inspiración te encuentre trabajando”, y por tanto, algo tiene que ver con un proceso de aproximaciones sucesivas, pero el cambio exponencial rompe la tendencia de valores de la cadena de iteraciones.

## **¿Somos seres sintácticos o semánticos?**

Llegados aquí, hagamos una reflexión sobre si en los seres humanos domina lo sintáctico frente a lo semántico, si en nuestra sociedad, en nuestra universidad domina lo sintáctico o lo semántico.

No todos los sistemas complejos con propiedades emergentes presentan inferencia semántica. Los ecosistemas naturales son un buen ejemplo de sistemas complejos plagados de propiedades emergentes y fenómenos de criticalidad, pero faltos de inferencia semántica.

Sin duda, un sistema donde todos sus elementos presentaran la propiedad estacionaria de ser semánticamente generativos daría como resultado una organización caótica donde no se daría tiempo a asentar los nuevos conceptos, alejándose dramáticamente de la explotación de su potencialidad. Es posible que por eso hayamos evolucionado de manera que sólo un número reducido de elementos de nuestra sociedad sean grandes generadores semánticos, popularmente denominados como genios. De hecho, las ideas geniales son fenómenos de muy baja frecuencia en los genios.

Son muy pocos elementos de nuestra sociedad a los que les podríamos otorgar el rango de genios, es incluso reducido el número a lo largo de toda la historia. Pero entonces, ¿qué somos los demás?, ¿seguidores de las genialidades?, ¿usuarios del conocimiento previamente establecido?, ¿elementos que sobreviven al devenir diario en un ciclo día-noche totalmente sintáctico? Es posible que no sea así, que quizá tengamos algunas ideas geniales, aunque su magnitud y frecuencia se limite a nuestro contexto. Recuerdo con gran “vividez” el caso de un paciente al que tenía que evaluar neuropsicológicamente con el objetivo de colaborar en el diagnóstico de demencia. Era una persona analfabeta y había trabajado toda su vida como agricultor en el campo. Preguntando a la familia por sus antecedentes me contaron una anécdota muy reveladora: “el sistema de acequias y regadíos de los campos del pueblo en el que vivían era sumamente ineficiente y el agua llegaba de manera muy desigual a diferentes partes de los campos de cultivo desde el río. Sin tener ningún conocimiento de la teoría de grafos y una casi inexistente educación, mi paciente, décadas atrás, fue capaz de crear un grafo donde los campos eran los nodos y los regadíos los enlaces. En un dibujo, que todavía conservaban en su casa, hizo el mapa de esos campos y regadíos e ideó la manera más eficiente de reestablecer la distribución del agua. Su idea más genial fue la de desarrollar un *hub* o nodo de alta densidad de conexiones que era el mejor distribuidor del agua. Esta idea surgió posiblemente de manera casi espontánea, sin basarse en un gran conocimiento previo. Euler, el matemático de la teoría de grafos, posiblemente tuvo una idea semejante muchos años antes, pero esta persona lo hizo casi de manera espontánea. De esta manera, podríamos decir que la gran mayoría de los seres humanos tenemos la potencialidad de tener ideas geniales que resuelven los problemas del día a día y que no están

plenamente basadas en conocimiento previo. La inferencia semántica es, por tanto, una propiedad intrínseca del sistema cognitivo del ser humano.

### **¿Cuál es entonces la frecuencia de aparición de la inferencia semántica en un humano medio?**

Un ejemplo de comportamiento extremadamente sintáctico son nuestras reglas de aprendizaje, que han permitido a los psicólogos entender la conducta humana y predecir determinados comportamientos. Cuando el profesor Luis Llavona me explicó, en mi época de estudiante en la UCM, las reglas del condicionamiento clásico y operante y las reglas de refuerzo y castigo (positivo y negativo), descubrí la enorme potencialidad de la psicología, pero al mismo tiempo me di cuenta de la simpleza del comportamiento que se podía reducir a unas reglas fundamentales. Lo más importante de aquella clase llegó cuando nos advirtió que no somos autómatas de reglas simples, las emociones nos introducen un nivel de complejidad que todavía no hemos podido formalizar al completo y por ello la investigación en psicología es tan apasionante.

Por tanto, podemos decir que el sistema cognitivo humano está impregnado de reglas sintácticas que se pueden llegar a formalizar y en ello se ha basado gran parte de la investigación empírica en psicología. Sin embargo, esto se combina con un sistema semántico al que hemos llegado a formalizar, de una manera todavía imprecisa, sus representaciones, pero no hemos explicado de una manera clara sus propiedades de inferencia semántica. Así, somos una mezcla entre sintáctica y semántica (formalizable y generativa).

Esto ha influido directamente en lo que hoy entendemos por IA. Un sistema inspirado en la parte que podemos formalizar de la cognición: sus reglas sintácticas y sus representaciones semánticas. Pero la IA no ha llegado a introducir todavía las emociones y la inferencia semántica, dos propiedades que nos hacen ser lo que somos y lo que nos ha llevado a tener éxito como especie.

¿Cuál sería entonces la diferencia entre ChatGPT y un humano? Al fin y al cabo, la IA la generaron los humanos, sus reglas y su conocimiento lo generaron los humanos. Siendo una IA sintáctica, poco se acerca a la inteligencia humana y por eso no debemos otorgarle decisiones que sólo nuestro sistema imperfecto debe tomar. Podríamos tener la tentación de que una IA gestionara los recursos energéticos, porque al estar exenta de sesgos emocionales y creatividades excesivas llegaría a soluciones sintácticas que optimizaría enormemente la distribución de

los recursos. No puedo imaginar una maximización sintáctica de la distribución energética. Posiblemente alguno de nosotros nos encontraríamos en la situación de dar argumentos de “humanidad” a una IA para que se le dé más temperatura a la casa de una persona con discapacidad.

## **IA y el debate abierto en la comunidad universitaria**

Con la llegada de la IA, ¿se perderán puestos de trabajo?, ¿podrá una IA sustituir a un médico?, ¿podría una IA sustituir al profesor universitario? Contestemos a esta última pregunta primero, por razones obvias. Si la enseñanza universitaria se basa en clases en las que la información que le damos a nuestros estudiantes es meramente sintáctica o es meramente una transmisión de un libro a un orador y del orador al estudiante, entonces somos perfectamente prescindibles y sin duda seríamos sustituidos antes o después. Ya existen en el mercado IA capaces de sustituir a ese tipo de profesor y además ni duermen, ni descansan. Pueden dar la misma clase a 4 grupos seguidos el mismo día y no acabar devastada. Sin embargo, si nuestra enseñanza esta impregnada por la semántica, por despertar el sentido crítico, por alentar la curiosidad del estudiante, entonces tenemos una oportunidad. Todos deberíamos reflexionar si las clases que damos a diario son sintácticas o semánticas. Más aún deberíamos reflexionar si la UCM es una universidad sintáctica o semántica, quizá sea en eso en lo que las instituciones se diferencien entre sí.

Debemos saber aprovechar lo que es intrínsecamente más humano. Un ejemplo de humanización en la Universidad Complutense son las clínicas de psicología y de logopedia de la Facultad de Psicología de la UCM, así como el servicio de PsicoCall. La clínica de psicología ha realizado más de 5.000 consultas el año pasado, en pacientes con diferentes patologías psicológicas tales como depresión, fobias, estrés, ansiedad, etc. La clínica de logopedia interviene en niños y adultos con trastornos del lenguaje debidos a diferentes causas. Además, se ofrece el servicio PsicoCall que atiende a estudiantes universitarios que necesitan un apoyo psicológico. Sin duda, estas clínicas ofrecen la atención especializada, de alto rigor científico, que requieren los pacientes, pero sobre todo ofrecen una atención humana para trastornos psíquicos que padecen seres humanos. Las IA no pueden sustituir la necesidad del ser humano de compartir sus dolencias

con otro ser humano y esperar un trato emocional del mismo calado. De esta manera, esta relación profesional y emocional entre humanos entra dentro del mundo semántico, no simulable completamente por una IA (ej., lenguaje no-verbal, emociones). Lo mismo diría de los servicios médicos de los que disfrutamos en el campus de Moncloa y en el de Somosaguas.

Recuerdo que le pregunté a mi buena amiga Gretchen Von Allmen, directora del departamento de Neurología Pediátrica del Hospital Hermann Hospital y catedrática de la Universidad de Houston, Texas: ¿cuál es la principal cualidad que debe tener un médico?; su respuesta fue contundente “ser capaz de entender el sufrimiento de los demás”, y yo le contesté “entonces tener teoría de la mente y empatía, es decir, ser capaz de sentir lo que sienten los demás”. Sin duda, es una capacidad que está íntimamente ligada a nuestras emociones. Son esas habilidades las que nos convierten en más humanos pensar en “nosotros” y colaborar entre “nosotros”.

Por tanto, está claro que los seres humanos nos necesitamos unos a otros, esa ha sido la clave de nuestra evolución como especie tal y como indica el famoso antropólogo Yuval Noah Harari en su libro Sapiens. Son diversos los ejemplos de que cuando un humano es aislado de otros se ha incurrido en el mínimo desarrollo de la capacidad del lenguaje, atención, función ejecutiva o percepción. Es más, no se vieron potenciadas las emociones y los individuos eran egoístas y tendían al aislamiento. De esta manera, este desarrollo emocional y cognitivo depende de que otros sistemas semánticos interactúen con nosotros.

Por tanto, el riesgo de un proceso de deshumanización sería el camino desde lo semántico a lo sintáctico y es ese el viaje que tenemos que regular y definir sus límites y necesidades reales. Pero debemos parar y pensar, predecir y matizar, tal y como pedía recientemente Geoffrey Hinton, uno de los padres de la IA, en un artículo en el New York Times. No debemos dejar que los acontecimientos nos abrumen, ni los hechos consumados nos determinen. Las IA pueden ser unas grandes aliadas para muchísimas cosas en nuestra sociedad y será seguro un gran beneficio para todos. Pero necesitamos entender mejor cómo definir los caminos por donde debería transitar.

Déjenme que finalice esta lección con una declaración: “nunca revelaré si este discurso fue escrito por chat GPT o realmente por un humano” prefiero dejarlo a juicio de la audiencia.

# Referencias

- Aguado, L. (2019). *Cuando la mente encontró a su cerebro*. Alianza Editorial.
- Anderson, J. R. (1995). *Cognitive psychology and its implications*. New York: Freeman.
- Baddeley, A. y Hitch, G. J. (1974). "Working memory". In G. Bower (ed.), *The psychology of learning and motivation* (pp. 47-89). New York: Academic Press.
- Baddeley, A. (1999). *Essentials of Human Memory*. Hove, UK: Psychology Press.
- Collins, A. M. y Quillian, M. (1969). "Retrieval time from semantic memory". *Journal of Verbal Learning and Verbal Behavior*, 8(2), 240-247.
- Damasio, A. R. (1996). "The somatic marker hypothesis and the possible functions of the prefrontal cortex". *Philos Trans R Soc Lond B Biol Sci*, 35 (1346), 1413-1420.
- Eichenbaum, H. (2002). *The cognitive Neuroscience of Memory: an introduction*. New York: Oxford University Press.
- Ekman, P.; Levenson, R. W. y Friesen, W. V. (1983). "Autonomic nervous system activity distinguishes among emotions". *Science* 221 (4616): 1208-1210.
- Friston, K. J. (2002). "Beyond phrenology". *Annuals Review in Neuroscience*, 25, 221-250.
- Fuster, J. M. (1995). *Memory in the cerebral cortex*. Cambridge MA: MIT Press.
- Fuster, J. M. (2003). *Cortex and Mind: Unifying Cognition*. New York: Oxford University Press.
- Graf, P. y Schacter, D. L. (1985). "Implicit and explicit memory for new associations in normal and amnesic subjects". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11 (3), 501-518.
- Hebb, D. O. (1949). *Organization of behavior*. New York: John Wiley y Sons.
- Johnson Laird, P. N. (1999). "Mental models". In R. A. Wilson y F. C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 525-527). Cambridge, MA: MIT Press.
- Kellogg, R. T. (2003). *Cognitive psychology* (Vol. XXII). Thousand Oaks, CA, US: Sage Publications.
- Marr, D. (1982). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA.
- Mehrabian, A. (1971). *Silent Messages* (1st edition). Belmont, CA: Wadsworth. ISBN 0-534-00910-7.
- McClelland, J. L. y Rumelhart, D. E. (1985). "Distributed memory and the representation of general and specific information". *Journal of Experimental Psychology: General*, 114 (2), 159-188.
- Neisser, U. (1967). *Cognitive psychology*. East Norwalk, CT, US: Appleton-Century-Crofts.

- Newell, A. y Simon, H. A. (1972). *Human problem solving*. Oxford, England: Prentice-Hall.
- Quillian, M. (1968). *Semantic Memory*.
- Rivière, Á. (1991). *Objetos con mente*. Madrid: Alianza.
- Rumelhart, D. E. (1989). "The architecture of mind: a conexionist approach". In I. Posner (ed.), *Foundations of Cognitive Science*. Cambridge MA: MIT Press.
- Ruiz-Vargas, J. M. (2002). *Memoria y olvido, perspectivas evolucionista, cognitiva y neuro-cognitiva*. Madrid: Editorial Trotta.
- Ruiz-Vargas (2023). *La memoria y la vida. Somos lo que recordamos* (ensayo y pensamiento). Editorial Debate.
- Shallice, T y Cooper, P.C. (2011). "The organisation of Mind". *Cortex*. 48 (10):1366-70.
- Schacter, D. L. y Tulving, E. (1994). "What are the memory systems of 1994?" In D. L. Schacter y Tulving (eds.), *Memory Systems* (pp.1-38). Cambridge, MA: MIT Press.
- Squire, L. R. (1986). "Mechanisms of memory". *Science*, 232 (4758), 1612-1619.
- Sternberg, R. J. (2003). *Cognitive Psychology*. Wadsworth, Thomson Learning.
- Tulving, E. (1972). "Episodic and semantic memory". In Tulving y W. Donaldson (eds.), *Organization of Memory* (pp. 381-403). New York: Academic Press.
- Tulving, E. (2002). "Episodic memory: from mind to brain". *Annual review of Psychology*, 53, 1-25.
- Tulving, E. y Craik, F. I. (2000). "Preface". In Tulving y F. I. Craik (eds.), *The Oxford Handbook of Memory* (pp. V-VII). Oxford: Oxford University Press.



