



UNIVERSIDAD DE CHILE
FACULTAD DE FILOSOFÍA Y HUMANIDADES
DEPARTAMENTO DE FILOSOFÍA
SEMINARIO DE FILOSOFÍA CONTEMPORÁNEA

El Conexionismo Naturalizado, Modelos de la Mente y Representación Mental

Informe de seminario de grado:

'La Mente en un Mundo Físico'

Profesor guía Guido Vallejos O.

Alumno Mauricio Toro E.

Marzo del 2009

Índice

Introducción	2
I Representación mental	4
I.I Mente como procesador de símbolos	7
I.II Mente como patrones de activación de nodos dentro de una red de procesadores.	9
II Conexionismo, o la historia del la computación naturalizada.	10
II.I Redes Neurales	13
II.I.I Ejemplos de redes neurales con comportamiento 'inteligente'	18
III Problemas asociados al conexionismo: sistematicidad y productividad de los procesos mentales.	27
Conclusión, un experimento mental	36
Bibliografía	39

Introducción

La filosofía de la mente ha tenido un impresionante desarrollo en los últimos años, aun cuando este no sea tan claro como el que ocurre en las ciencias naturales, donde este puede verse en su poder comprensivo sobre los fenómenos del mundo, es decir, el poder explicativo y predictivo de sus teorías. Si bien el desarrollo de la filosofía de la mente tiene que ver con mayor comprensión sobre los fenómenos mentales -hay tantas formas de referirnos a la mente, sus propiedades internas y sus relaciones que no puede ser que una de ellas no sea algo más correcta que otras- no parece haber un consenso sobre qué teoría es más o menos correcta. Todo esto se refleja en que hay más y mejores explicaciones, junto con más y mejores predicciones para los fenómenos mentales. El problema es que estas también son tantas y tan distintas entre ellas que no se hace difícil reconocer a las que tienen algo de verdad. Pero, es posible determinar un par de condiciones que debe cumplir una teoría de la mente para ser tomada en serio y merecer fondos públicos, atención de los medios, cursos en las universidades y una estrella en la frente. Estas tienen relación con el mencionado desarrollo de esta disciplina, avalado por las mayores capacidades de las teorías, sus compromisos ontológicos, metodológicos, epistemológicos, en el fondo, todo el tejemaneje filosófico y la coherencia de todo esto con lo que se puede decir del mundo con la ayuda de la ciencia.

Entre las diversas teorías, debemos poner en un lugar privilegiado a las que, al menos, permitan hablar de la mente desde proposiciones verificables, aun cuando esto no sea factible por deficiencias en las capacidades técnicas o una neurociencia joven, que se haga cargo de la relación entre la mente y el mundo y, por último, que permita explicar el contenido mental y las relaciones entre las distintas propiedades mentales. De manera más precisa: Las teorías que permitan explicar la mente en relación a su contenido, dando cuenta de cómo se producen y relacionan los estados mentales y la relación de la mente con el mundo. Todo esto manteniendo la coherencia de esa teoría con lo que nos digan las distintas ciencias cognitivas, aquellas que pretenden dar cuenta de los hechos mentales como propiedades del cerebro y del sistema nervioso central.

Dentro del subconjunto de teorías que he acotado, me detendré en aquellas que sostienen que la mente, y toda la 'mochila cognitiva', puede y debe ser explicada en términos de fenómenos naturales, que los fenómenos mentales tienen partes constituyentes y que la transformación/almacenamiento de

esas partes es lo que los constituye. Pues estas teorías parecen tener las distintas características que permiten hablar de avance dentro de una disciplina, estas por una parte se comprometen con maneras de explicitar sus compromisos teóricos con modelos funcionales implementables en computadores, de manera que sus predicciones son evaluables, además que se sostienen sobre una comprensión materialista de la mente, relevante para su consistencia con las ciencias y la neurociencia en particular.

De todas las teorías que podrían describirse aquí, me remitiré a hablar sobre aquellas que se construyen sobre las representaciones mentales, pues en este respecto hay bastantes diferencias de opinión sobre sus propiedades y las características propias de esta clase de fenómenos, con vistas a encontrar cual es la mejor manera de dar cuenta de ellas y los modelos que están más cerca de una explicación que nos permita expresarlas desde una perspectiva materialista. Sobre esto mucho se ha dicho, no pretendo ser el último en hablar al respecto, ni tampoco zanjar el problema sobre la manera correcta de hablar sobre ellas. La modesta finalidad de éste texto es presentar una aproximación que es a primeras luces contraintuitiva, pero que pretende explicar no sólo la relación entre nuestras representaciones mentales y la psicología, sino al mismo tiempo dar cuenta de cómo las funciones del cerebro están involucradas en este proceso. Esta, por ser parte del materialismo naturalista, tiene la ventaja del compromiso y coherencia con el materialismo y el discurso de las ciencias naturales, será comparada con una aproximación que busca primero responder a las necesidades impuestas por la psicología y la descripción tradicional de los fenómenos mentales.

Esto pues me parece que la primera tiene una explicación más completa, teniendo en cuenta no solo las constricciones impuestas por la psicología, sino que antes de ellas las impuestas por la biología y las neurociencias, tales constricciones pueden caracterizarse como:

- Dar cuenta de la representaciones como un producto del cerebro, comprender la mente y sus procesos como un fenómeno natural.

- Explicar la mente y el cerebro como productos del desarrollo evolutivo, la complejidad del cerebro y su desarrollo en la historia evolutiva de los seres vivos implica que este plano sea considerado, una teoría que no sea capaz de explicar la relación entre el modo de producir representaciones propio de los humanos y el de los animales en términos de desarrollo evolutivo del cerebro debe ser descartado.

- La relación entre representación y lenguaje, el hecho que las propiedades mentales tengan un historial evolutivo hace que las representaciones mentales sean anteriores al lenguaje y que nosotros los

humanos seamos capaces de expresarlas y ordenarlas en esta manera, mediante una gramática, no implica que sea la forma en que ellas se dan. Pero si hace necesario aclarar cómo se complementan estas dos formas de representación, con esto se ayudaría también a la comprensión del desarrollo cultural, pues a medida que el lenguaje se ha complejizado y podemos comunicar nuestras experiencias y predicciones del mundo de manera más clara, somos capaces de comprender más partes del mundo y con ello representárnoslo de manera más completa.

Las teorías de la representación que he escogido pretenden llegar a modelos implementables en computadores, estos pretenden tener características, o más bien comportamiento, similares a las de nuestra mente. Por lo que estas constricciones deben complementarse con las que implica el modelar en un tipo de arquitectura distinta a la del cerebro, de manera que esos modelos sean capaces de sobrellevar las limitaciones que implica esta diferencia. Entre las que pueden mencionarse está el hecho que, por ejemplo, las neuronas funcionan de manera radicalmente distinta a los chips de una computadora; la memoria, en términos naturales, no funciona como un disco duro; y, por último, que el procesamiento de información en el cerebro es paralelo, a diferencia de la forma serial en que lo hacen las computadoras¹.

Este informe está lejos de presentar una tesis y defenderla, de hecho es más bien una manera de dar cuenta del estado de la discusión sobre la teoría de la representación y los modelos mentales más comentados. Se presentan los modelos que pretenden cumplir con las características de la mente, y estos son sopesados con nuestro conocimiento sobre las propiedades del cerebro y la luz que la psicología puede aportar sobre nuestra vida mental. La conclusión que se presenta, en realidad, no es tal, más bien es un posible problema para desarrollar a la larga y pretende promover la discusión más que cerrarla.

I Representación mental

Primero es necesario aclarar qué quiere decir aquello de representación mental, de aquí en adelante RM. Desde que se asume que los fenómenos mentales tienen partes constituyentes y que el contenido, función y cuanto pueda decirse de ellos depende de estas, a esas partes constituyentes se les

¹ Para una mayor profundización en las constricciones empíricas que debe cumplir una teoría de la representación se puede revisar en Churchland, P.S. (2002), en particular el capítulo 7.

bautizó como RM².

Nuestra experiencia del mundo, o más bien la experiencia de nosotros en él, se produce en la interacción mutua. Nuestros sentidos enfrentados a distintos estímulos y los mecanismos de nuestro cerebro procesándolos producen nuestra vida mental. El hecho mental de percibir una flor, calcular una suma o escuchar un sonido, se nos presenta como un compuesto, distintas estructuras relacionándose entre ellas, implicadas por esa interacción con el mundo o al menos con lo que hemos aprendido de él. Cada una de ellas se relaciona con una o más, haciendo de esa flor que vemos una rosa o una amapola o de ese sonido un do o un fa, todo esto gracias a la manera consistente, coherente y persistente de relacionarse que tienen estas. Estas se organizan y funcionan de manera similar a relaciones entre símbolos con propiedades semánticas

Asumir la existencia de las RM implica creer que lo que hacemos en nuestra mente es formar algo como un modelo del mundo, ordenamos el conjunto de experiencias que tenemos según las partes que la componen y relacionamos estas partes según la manera en que entendemos que se organizan. El que asumamos una postura materialista implica que estas estructuras se producen en el cerebro, una determinada estructura mental esta relacionada con una configuración particular del cerebro.

Como se supone que los procesos mentales son relaciones entre RM, se debe asumir que esas maneras de relacionarse deben respetar algún tipo de reglas y que las RM deben tener algún tipo de regla de conformación, es decir, que los procesos mentales son algún tipo de computación, donde las RM serían los objetos computados. Según esto, la mente funcionaría de alguna manera parecida a la que lo hace un computador y el tipo de computación que esta realiza es lo que marcará la diferencia entre las distintas teorías de la mente. Mientras unos consideran que el cerebro es un procesador de estructuras simbólicas atómicas con contenido semántico, las RM, y que los procesos mentales son manipulaciones de estas según reglas que dependen de su estructura constitutiva, hay otros que afirman que las RM son resultado de activación de patrones de una red con múltiples procesadores y los procesos mentales consisten en la activación dispersa de esos patrones³. La manera de computar las

2 Si bien hay casos de filósofos que no consideran que los fenómenos mentales tengan una estructura y partes constituyentes, según parece -y pretendo de alguna manera mostrar en este ensayo- es necesario postular estas entidades para dar cuenta de lo que ocurre en la mente, al menos mientras no tengamos una mejor manera de explicar la diversidad del contenido mental.

3 Estoy reduciendo el campo de ambas teorías, pues hay defensores de las RM simbólicas que no se convencen de la dependencia entre las reglas del procesamiento mental y la máquina en que se implementan y hay defensores de las RM en redes que no las consideran distribuidas. No he tomado todos los casos por un problema de espacio y por las

RM define el método explicativo de la vida psicológica de los sujetos, la manera de generar predicciones sobre la conducta de estos, es decir, cómo se relacionan las RM define la explicación que se puede dar al contenido y a las propiedades causales de los fenómenos mentales, además la diferencia en la forma de computar el contenido mental determina qué tipo de modelo computacional se propondrá. La coherencia de todo esto, la manera de computar las RM, la forma en que se producen y cómo se relacionan junto con sus poderes causales, con lo que ocurre en cerebro y la coherencia de esta explicación con la ciencia es lo que definirá, en última instancia, qué teoría merece la ya mencionada estrella en la frente.

Cuando tratamos de dar una generalización que permita dar cuenta de la conducta de un sujeto o del contenido de un estado mental propio, lo que hacemos es expresarlas en términos de la relación intencional en la que se está respecto a una proposición, se *crea* tal, *desea* cual, *sabe* anga o *teme* manga, estas son llamadas actitudes proposicionales, la idea de intencionalidad tiene relación con la idea de *ser-respecto-a*, *aboutness*. Esto ha implicado que muchos postulen que estas actitudes proposicionales son parte necesaria del discurso de cualquier teoría sobre RM, se exige que una explicación de los fenómenos mentales que los considere como organizaciones de RM, postule a estos como algún tipo relación de símbolos con contenido intencional. Por otra parte, se han presentado algunos con la idea de que esta manera de realizar generalizaciones no es tan buena como pretende y que el hecho de ser parte de nuestra tradición explicativa no la hace mejor. Esto fundamentado en el escaso poder explicativo de esta aproximación y lo lejos que se encuentra de ser consistente con las ciencias naturales, pues la manera en que pretenden explicar la computación de las RM no es homologable a ninguna de las situaciones que se pueden apreciar en el funcionamiento del cerebro.

La postura que se tome respecto a la naturaleza de las RM y a la intencionalidad de lo mental es una de las características que define qué clase de computación realiza la mente. Si se asume que las RM son estructuras atómicas con contenido intencional, se asume que la mente realiza algún tipo de algoritmos con las RM, manipulando símbolos con contenido intencional según sus valores y regulados por las propiedades de ese contenido y esa estructura. Es decir, cada estructura implica un algoritmo particular y se relaciona con otras según algoritmos determinados, hay una manera particular en la que interactúan entre sí las distintas RM, hay un lenguaje del pensamiento que regula cómo funciona

consideraciones ya comentadas sobre la necesidad de un modelo que sea materialista, para una historia más completa de las RM y sus distintos sabores se puede revisar la entrada al respecto en Wilson & Klei (1999).

nuestra mente. Por otra parte, asumir que la hipótesis de las actitudes proposicionales y las implicadas por la psicología *folk* son falsas, si bien no automáticamente, se relaciona con la concepción de la mente como una compleja red que realiza múltiples procesos de manera simultánea y que la instanciación de una determinada RM depende del procesamiento paralelo de múltiples algoritmos similares. El problema de esto es que se hace necesario aclarar cómo pueden *ser-respecto-a* algo, si no son estructuras atómicas. Una forma de salir de este problema es apelar a una computación sub-simbólica, que implicaría que estas propiedades serían emergentes de un nivel de computación más simple.

Teniendo en cuenta esto, paso a relatar un poco la historia que cuentan las dos maneras de explicar el procesamiento de información dentro de la mente antes mencionadas, para dar una primera aproximación a las distintas maneras de enfrentar el problema de modelos mentales e inteligencia artificial⁴.

I.I Mente como procesador de símbolos

(mente/cerebro~software/hardware, clasicismo)

La primera aproximación a una descripción del procesamiento de las RM asume que estas son estructuras que contienen información del tipo que se hace referencia en la psicología *folk*, aquella que describe la vida mental en términos de deseos, creencias, conceptos y demás. Estas estructuras son símbolos que contienen información y se relacionan entre ellas según leyes de sintaxis que le son propias, donde su significado está determinado por la manera en que se producen y/o sostienen. Esto quiere decir que las leyes que rigen sus poderes causales, respecto a otras estructuras mentales o bien respecto a conductas en el mundo, dependen de la manera en que se relacionan/producen unas y otras estructuras. Este conjunto de reglas y estructuras funciona como un lenguaje, de hecho a esta hipótesis se le llama lenguaje del pensamiento (LDP), respetando las reglas de la lógica clásica y es implementable en una computadora serial. Las distintas frases en este lenguaje, al relacionarse entre sí,

⁴ Las caracterizaciones son una recopilación de las descripciones de ambas teorías tomadas de Smolensky (1989) y Fodor (1985).

implican o bien otras estructuras o frases o bien conductas.

La relación que se supone hay entre esta descripción y el cuento de la ciencia, está en que esos símbolos se producirían en el cerebro, una determinada neurona o un grupo de ellas se activaría al estar procesando determinado símbolo o realizando un determinado proceso con un enunciado. Las distintas partes del cerebro interactúan entre ellas de manera similar a la que lo hacen los enunciados en este lenguaje del pensamiento, en un sentido muy amplio estaríamos hechos de manera que nuestro procesamiento de información se da de forma lógica, hemos desarrollado nuestra racionalidad producto de la evolución, aún cuando llegar a estas conclusiones es llevar el argumento un tanto más lejos de lo que pretende. Decir que nuestro cerebro y sus procesos funcionan de manera similar a como funciona nuestra mente y los procesos propios de ella, es suficiente.

Si bien no todos los realistas intencionales, aquellos que defienden que la descripción de la vida mental debe hacerse en los términos de la psicología *folk* y que estos son reales, están de acuerdo con esta hipótesis, la del LDP, afirmar que los procesos mentales tienen contenido intencional y que las actitudes proposicionales son la mejor manera de dar cuenta de estos, implica asumir que estos tienen una semántica propia, donde los objetos intencionales se relacionan entre sí y pueden aparecer en distintas partes de una RM. En cuanto estos determinan la conducta y esta parece tener una organización interna y elementos recurrentes, si los procesos mentales no funcionasen de una manera similar, con una coherencia interna y elementos recurrentes e intercambiables, habría algo oliendo mal en nuestra explicación.

Tenemos, entonces, dentro de la cabeza un tipo de estructuras con contenido intencional, que se relacionan entre sí, las cuales según sus propiedades internas son manipuladas por procesos neurofisiológicos regulares y esto implicaría el que se produzca una determinada respuesta conductual.

Dentro de las ventajas de esta aproximación están las implicaciones que tiene dentro del desarrollo de inteligencia artificial, pues teniendo una manera de expresar los procesos mentales de manera que estos aparecen como homólogos a procesos computacionales, es más fácil imaginar una manera de crear programas capaces de emular estos algoritmos para la manipulación de información de modo que el comportamiento de este sea similar al que tendría un ser humano. Además de dar la posibilidad de descubrir las leyes que regulan las maneras de generar explicaciones y predicciones que

usamos en el cotidiano.

I.II Mente como patrones de activación de nodos dentro de una red de procesadores.

(mente/cerebro=ware, computación sucia 'yo-relevante', conexionismo naturalizado)

Si, por otra parte, consideramos que los procesos mentales no deben ser explicados, necesariamente, en términos que den cuenta de la manera en que tradicionalmente los expresamos, sino que la explicación debe ser coherente con la manera en que está organizado el cerebro y la manera en que este funciona, esto es que los procesos mentales son procesos cerebrales y que, por lo mismo, la explicación sobre la manera en que están estructurados, junto a la forma en que se relacionan, se debe hacer dando cuenta de cómo está estructurado el cerebro y cómo se relaciona internamente. Estamos en manos de los conexionistas, quienes afirman que la mente/cerebro es una red de procesadores conectados entre sí, cuyos patrones de activación son los que determinan el contenido de nuestras experiencias mentales.

La explicación de los fenómenos mentales y la predicción de conductas según estos modelos se relaciona con los patrones de activación que ocurren en los nodos de la red, luego, las propiedades internas de los fenómenos mentales, es decir el contenido de las RM, está definido por ellos. Si bien se puede hablar de contenido semántico respecto a la RM o dentro de un fenómeno mental, ni los nodos, ni sus patrones de activación, considerados individualmente lo tienen. La manera en que los distintos nodos manipulan y relacionan la información que reciben y envían es lo que hace que se den tales o cuales RM o que se realice tal o cual conducta. La fuerza con que están conectados entre sí los distintos nodos depende del entrenamiento que ha tenido la red en su desarrollo y la distribución de estas conexiones es lo que determina la manera en que se comunican entre sí y, en última instancia, es esta distribución lo que determina el patrón de activación de los nodos. La comprensión del modo en que se procesa la información en las distintas estructuras constitutivas de la red, además de las propiedades internas de sus nodos constituyentes, junto con definir la situación en la que se encuentra la red en un determinado momento, permitirían formular las explicaciones que eran buscadas. En cuanto las RM serían el patrón de activación dentro de una red neuronal, la relación causal entre RM y conducta estaría asegurada, pues parece de perogrullo que si un patrón puede activar las zonas relacionadas con producir la imagen de un billete en el piso, esta puede implicar que se active la zona que se relaciona

con mover mis músculos de manera que lo recoja.

Entre las ventajas de esta aproximación están: que es la más cercana al modo en que funciona el cerebro, esto era parte de los criterios relevantes al momento de evaluar una teoría; que da cuenta no sólo de los procesos mentales de individuos *normales*, si no que permite explicar las experiencias y las conductas anómalas en términos de deterioro o ausencia de nodos dentro del entramado; además de comprender la flexibilidad de los procesos cognitivos frente a situaciones novedosas, pues una nueva experiencia nunca es del todo nueva⁵.

II Conexionismo, o la historia del la computación naturalizada.

Como ya se ha comentado, es parte del imaginario de los filósofos de la mente, psicólogos y en general de cuanto interesado en el tema que los fenómenos mentales, junto con los procesos que estos implican, están constituidos por estructuras más simples que contienen la información que aparece en los primeros, que los fenómenos y procesos mentales son la transformación, almacenamiento y combinación de elementos en las estructuras. Las RM son parte de la descripción de casi cualquier filosofía de la mente desde tiempos del buen Aristóteles, pero su aceptación general no implica que haya acuerdo sobre su naturaleza, función y, en general, cuanto pueda decirse de ellos después de asumirlos como parte de los elementos que debe haber en una explicación. Incluso desde la aventajada posición de los materialistas, aventajada pues tienen siempre esa simpática aliada que son las ciencias naturales, tampoco hay mucho acuerdo, pues si bien asumen que la mente y los fenómenos que esta implica son de algún modo causados por (en) el cerebro, al hablar sobre las RM que los constituirían, se producen algunos quiebres, por un lado están los que buscan una explicación que sea coherente con la manera en que se nos hacen presentes los fenómenos mentales y la forma en que tradicionalmente los comunicamos, y en otro lado están los que dicen que la explicación debe ser buscada junto con las explicaciones que se pretenden dar sobre el funcionamiento de la maquinaria en la que se sustentan. Una breve pincelada sobre esto se dio en los apartados anteriores, ahora es tiempo de presentar al que me parece el candidato a ganador y los motivos para esto.

Para hacer claro mi *bias*: me centraré en contar la historia de las RM desde la descripción

⁵ Por ahí un filósofo decía que por muy extrañamente que se pintase un cuadro, los colores en él siempre nos serían reconocibles. Y si no lo ha dicho alguien, me parece que no es una idea que requiera mayor explicación.

materialista de los fenómenos mentales, naturalista y conexionista sobre las RM, que tiene como condiciones de evaluación sobre las hipótesis que postula no solo la introspección y coherencia con la tradición filosófica, sino que incluye la coherencia con la descripción científica del mundo, exige datos testeables y conclusiones observables. La hipótesis conexionista no es del todo nueva, igual que no es nuevo el acercamiento naturalista a la filosofía de la mente⁶, pero si son nuevos los modelos de redes neurales y las distintas aplicaciones que han tenido. Y me parece que han mostrado ser la manera metodológicamente correcta de proceder y esto se evidencia en la preocupación por explicar los fenómenos mentales tomando como partida el sustrato en que se realizan, además de las ventajas a la hora de explicar y solucionar problemas dentro de la psicología que tiene este tipo de modelos y que da pie para el desarrollo e implementación de estos en arquitecturas no biológicas con resultados altamente interesantes.

Comenzaré con describir lo más detalladamente que pueda qué son y cómo funcionan las redes neurales y cómo es que se producen las RM dentro de estas, junto con un par de ejemplos de cómo han mostrado que funcionan los modelos que se han desarrollado para implementar estas teorías.

Respecto a qué sean las RM en este sentido, estas deben ser comprendidas como la configuración de un patrón de activación dentro de una red de procesadores de información. Es claro que esta descripción requiere una aclaración. Por ahí dicen que lo mejor es partir por el principio, por lo que primero se explicará qué serían y cómo funcionan esos procesadores de información que se activan y comunican de maneras extrañas entre sí. Como estamos dentro de un modelo naturalista, debemos asumir que los procesos mentales son procesos dentro del cerebro, ejecutados por sus partes constituyentes, las neuronas. Estas minúsculas células serían las encargadas de ejecutar cualquier tipo de proceso del cerebro, poseen distintos elementos constituyentes y entre sí comparten distintos neurotransmisores, hormonas, impulsos eléctricos, cartas de recomendación y certificados de antecedentes, que serían las maneras de comunicarse entre sí. Un patrón de activación dentro de la red de neuronas sería la descripción de la configuración de las neuronas que están activadas en un determinado momento, cual es su nivel de excitación, los neurotransmisores que están compartiendo, qué señal, desde donde y hacia donde se está enviando, y toda la información que se pueda determinar

6 Un conocido médico griego, Hipócrates, el del juramento hipocrático, postulaba que nuestras experiencias sensibles y las emociones son reguladas en el cerebro, todo esto en tiempos cercanos a Aristóteles, pero sus palabras no fueron muy tomadas en serio.

y considerar relevante en un determinado tiempo constituyen ese patrón. Más adelante dedico un espacio a la descripción más completa de una neurona y con eso se aclara un poco el concepto.

Si pensamos que el abanico de sensaciones, emociones, olores, colores y sabores que tiene nuestra experiencia son de algún modo estados del cerebro, ¿cómo es que se debe explicar la manera en que estos se producen? ¿de qué manera está ordenada la información que hay en los hechos mentales que percibo? El hecho que tenemos acceso privilegiado a nuestra vida mental no implica que esto nos dé un conocimiento privilegiado de cómo es que esta se produce y las leyes que la rigen, pero somos porfiados y conservadores al momento de explicarnos nuestra conducta, creemos saber cómo es que esta se origina: *Hacemos algo porque queremos conseguir tal otra cosa*. Ninguno de nosotros piensa que cuando tomamos un vaso para tomar agua porque tenemos sed, en realidad lo que ocurre es que queremos tomar agua porque el nivel de hidratación de nuestro organismo está disminuyendo y esto hace que parte de las células que procesan la información interna sobre nuestro cuerpo empiecen a enviar información sobre esta situación a determinadas áreas del cerebro, las que codifican esta información y la comparten con otras áreas con las que está en contacto, haciendo que estas empiecen a compartir esta información como propia y se produce una reacción en cadena que culmina en la activación de una determinada configuración dentro del sistema que puede explicarse como: el sistema *tiene sed*. Aún menos claro es que esta configuración implica la activación de mecanismos que van desde el cerebro hacia los ojos, moviendo la retina hasta que se enfoca un vaso, donde luego esto se traduce en la posición y distancia relativa del objeto con respecto a los apéndices superiores, esta información es enviada a otra área que mueve un determinado apéndice en una determinada dirección, con una velocidad y ángulo controlados, se sujeta el objeto... Bueno, la idea se comprende. Otra cosa sería ponernos a hablar de estar triste por la situación económica de alguna isla de la polinesia, describir un fenómeno tal en esos términos parece algo imposible y francamente dudo que esa clase de descripciones para tal fenómeno tenga algún valor. Por esto mismo, no pretendo que sea necesario explicar toda la psicología en estos términos, pero es claro que comprender cómo se produce una conducta en esos términos permite el desarrollo de remedios y tratamientos para enfermedades que pueden alterar la conducta de un individuo, desde el mal de Párkinson hasta la depresión, y comprender cómo es que ocurren las experiencias anómalas en algunos sujetos, sinestesia, miembros fantasma o esas locas experiencias que recuerdo de mi última reunión con mis amigos chamanes yaqui.

A manera de resumen y para aclarar algunos puntos, lo que estoy diciendo es que la noción de

RM dentro de la descripción conexionista-naturalista se caracterizan por ser producto del procesamiento paralelo y distribuido de información dentro de una red de procesadores. El contenido semántico de una RM no está en las unidades de procesamiento más cercanas al *input* ni en el *output*, sino que da según el patrón de activación que ese *input* implica, lo que posiciona a esa respuesta en un vector determinado en el espacio de activación de la red, un espacio ordenado y multidimensional. Dentro de esa capa escondida, entre el *input* y el *output*, la información no es, necesariamente, almacenada de manera atómica, no se generan organizaciones únicas que se entregan como respuesta a un *input* determinado solamente, sino que se crean espacios de representación, con tantas dimensiones como unidades de procesamiento tenga esta capa. Una RM es descompuesta y almacenada en distintas partes dentro de cada uno de esos espacios, donde el espacio que se le asigna tiene relación con la activación que produce en cada nodo y su posición en el espacio de representaciones está determinada por el vector de activación que produce dentro de toda la red. El hecho que las representaciones estén distribuidas dentro de este espacio y que los nodos procesen la información de manera paralela, hace que la red sea capaz de relacionar de manera correcta las representaciones, asociar un *input* con un *output* determinado de manera recurrente y no arbitraria, aun cuando parte de los nodos que la conforman no estén funcionando de manera correcta o se hayan perdido parte de las conexiones⁷.

II.I Redes Neurales

Tenemos entonces un grupo de gente que quiere explicar los fenómenos mentales, los cree hechos complejos compuestos de estructuras más simples, las RM, junto con creer que la mente esta en directa relación con el cerebro y por lo mismo pretende que la explicación sea coherente no solo con la psicología y la filosofía, sino que también con las ciencias naturales. Es decir, la explicación no está completa con sólo tener un sistema filosófico que dé cuenta de los fenómenos mentales y de sus RM constituyentes, tampoco queda completo con agregarle la descripción de la relevancia de esos fenómenos para explicar el comportamiento de un individuo, sino que además debe poder dar cuenta de cómo es que se ha desarrollado esta posibilidad, cómo se producen y organizan esas estructuras en términos materiales y de qué manera se relacionan los fenómenos mentales con los conductuales.

El desarrollo tecnológico, de la mano con el científico, ha entregado información sobre cómo

⁷ Aclaraciones sobre los conceptos de representación distribuida y procesamiento paralelo pueden verse en Smolensky(1989), Churchland (1996) cap 2 y en Frith (2007) cap 5.

está formado y cómo funciona el cerebro, la manera en que está organizado a nivel molecular y celular, esto permite comprender cómo está organizado a nivel estructural. El grupo de investigadores tiene ahora formas de explicar el cerebro, cómo se desarrolla, las funciones asociadas a las distintas partes y cómo éstas están relacionadas. Además de esto, el desarrollo tecnológico les ha permitido implementar los modelos que implican sus teorías en máquinas y en este punto el grupo de investigadores ha decidido que es mejor comenzar a modelar tomando en cuenta la manera en que funciona el cerebro y desde esos modelos y esa línea de investigación buscar una teoría que explique la manera en que funcionan los fenómenos mentales.

El estudio del cerebro ha mostrado que éste se compone principalmente de dos tipos de células, neuronas y células gliales, donde las primeras funcionan como un conjunto de células discretas, interconectadas e interdependientes, las cuales reciben *inputs*, que son procesados en sus partes constituyentes y comparten sus *outputs* con las otras células con las que están conectadas. En el fondo, el cerebro sería una red de procesadores cuyas computaciones serían independientes entre sí, pero sus resultados son compartidos con aquellas a las que están conectadas. A esto se le llamó la doctrina de la neurona, en principio se consideró que las células gliales funcionaban como tejido conectivo, de ahí el nombre, que sólo mantenían a las neuronas en su posición y les entregaban nutrientes para mantenerlas en funcionamiento. En la actualidad se ha mostrado evidencia de la relevancia de las células gliales en los procesos cerebrales y se ha comenzado a estudiar más su importancia en ellos, pero aún se asume que las células más relevantes en el procesamiento de información son las neuronas.

Aclarado el hecho de que el cerebro funciona como una red de procesadores de información, queda por aclarar cómo es que estos funcionan. Para luego presentar un par de ejemplos de modelos de esto que son implementables en una computadora.

Una neurona es una célula con la capacidad de enviar y recibir señales, las que pueden resultar inhibitorias o excitatorias dependiendo del tipo de neurona y el estado en que se encuentre. Consta, por lo general, de 3 partes constituyentes: el cuerpo o soma, dendritas y un axón, los dos elementos finales son los encargados de establecer las relaciones sinápticas, en la cual las prolongaciones, dendritas, que conectan una neurona con las terminales de otras, reciben las señales emitidas por las demás neuronas a la que está conectada en pequeñas espinas que están a lo largo de su extensión. Las señales son procesadas en el cuerpo de la célula y dependiendo del resultado de la suma del valor de éstas,

responde enviando o no una señal a lo largo de su axón a las neuronas con las que está en conexión (Fig. 1).

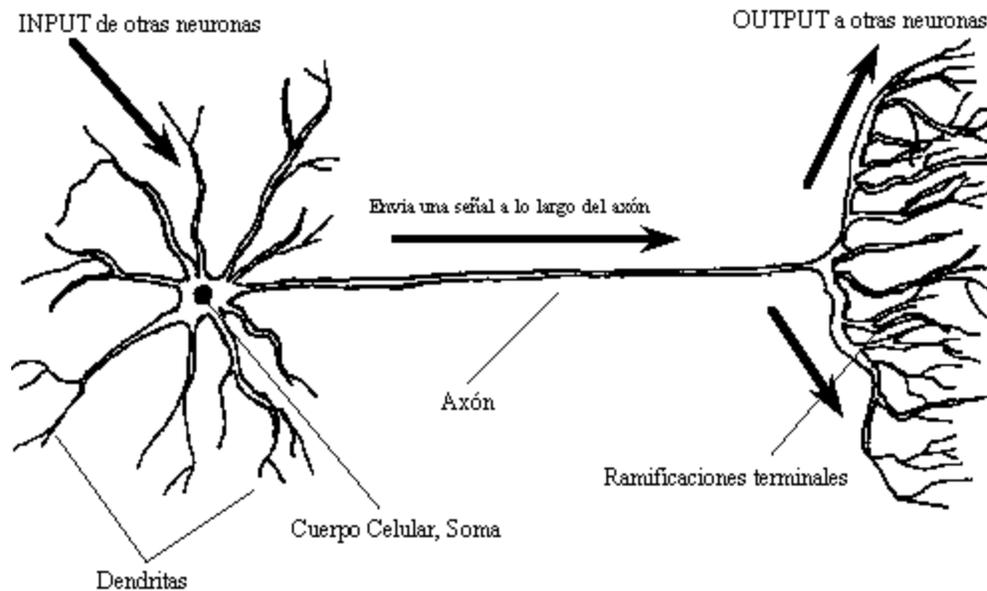


Figura 1 Dibujo de una neurona simple, tomado de Chad Lorden, 1996.

En el cerebro las conexiones, sinapsis, entre neuronas no son directas, las ramificaciones terminales de una neurona no tocan las dendritas de otra, al final de ellas hay pequeños botones que contienen los elementos químicos necesarios para la comunicación entre una neurona y otra. Los botones sinápticos de una están separados por una distancia de entre 20 a 40 nm de las dendritas de las otras y en este espacio es que tiene lugar la comunicación química, la neurona pre-sináptica libera neurotransmisores que son recibidos por la post-sináptica. Hay casos en que el espacio sináptico es tan pequeño, menos de 5 nm, que la neurona pre sináptica puede comunicarse enviando impulsos eléctricos, esto es una comunicación mucho más veloz y más confiable, pero son menos plásticas y menos comunes que las químicas. Los neurotransmisores son los que definen el carácter excitatorio o inhibitorio de la señal, dependiendo de el efecto que tengan sobre la neurona que las recibe.

La respuesta a la suma de las señales excitatorias o inhibitorias que da la neurona es todo-o-nada, esto quiere decir que o bien la neurona responde enviando su señal a todas sus terminales o a ninguna. Junto con el todo-o-nada, está el hecho que la neurona envía siempre un estímulo con la misma intensidad, lo que puede variar es la cantidad de respuestas que entrega en un tiempo, la frecuencia de disparo de la señal. Si consideramos que los *inputs* excitatorios aumentan la posibilidad de respuesta de

la neurona y los inhibitorios la disminuyen, podemos imaginar a los primeros como un valor positivo y los segundos como negativo. Si la suma del total de estas señales supera un determinado umbral, la neurona responde enviando una señal a lo largo de todo su axón y el valor de la suma determina la frecuencia con que enviará esa señal.

El conocimiento sobre las estructuras constituyentes de los niveles más básicos del cerebro permite un análisis más profundo sobre el funcionamiento de él, dando paso a una comprensión del cerebro en cuanto a distintas áreas funcionando de manera coherente y ordenada. Las distintas partes que componen el cerebro presentan una correspondencia anato-histológica en los mamíferos, esto es, las distintas áreas que componen el cerebro son comunes a todos los mamíferos, lo que varía es la manera en que éstas se relacionan entre sí, cuan interconectadas están unas con otras, además de los tamaños relativos de cada una de estas partes. Pero aun cuando entre miembros de una misma especie los tamaños de las distintas áreas del cerebro pueden variar en gran medida, perciben el mismo mundo. Del mismo modo, entre los miembros del reino animal compartimos también una manera de funcionar similar: si tiene sistema nervioso, tiene neuronas y éstas se relacionan de manera análoga a cómo lo hacen las nuestras. Las distintas áreas que componen el SN de cualquier ser vivo se componen de capas interconectadas, donde las capas inferiores proyectan sus *outputs* a las superiores y en algunos casos reciben *inputs* de las superiores también. Otra característica interesante del cerebro es el hecho que funciona procesando información de manera paralela, un grupo de neuronas puede estar tratando un problema, mientras otro puede estar trabajando en otro distinto, pero esto ocurriendo al mismo tiempo, esto permite el procesamiento de información de esta manera sea más veloz que hacerlo en una computadora serial, que computa mediante la operación de símbolos discretos de manera secuencial, a esta característica se le llama procesamiento paralelo distribuido (PDP).

El conocimiento de todo esto ha permitido postular modelos que representan lo más fielmente posible esta manera de procesar información, a estos se les ha llamado redes neurales. Estas consisten en distintos nodos, neuronodos o procesadores, distribuidos en distintos niveles y conectados entre sí. En el nivel inferior se encuentran las unidades que representan el estímulo externo, estas son la información que recibe la red y que es comunicada a la capa siguiente, está procesa estos valores y comparten el resultado de ese proceso a un tercer nivel que puede volver a procesar estos valores y así hasta generar un *output* (Fig. 2). Tomemos por ejemplo una red entrenada para el reconocimiento de

rostros: si el *input* que recibe es una imagen, el *output* puede ser el nombre, género y datos del sujeto presentado, en el caso que este sea parte del grupo de los rostros en el conjunto de entrenamiento. En caso contrario, el *output* pueden ser los datos relevantes que puede reconocer en una instancia novedosa, como el género de la persona o diferenciar entre una persona y objetos, dependiendo qué tan bien entrenada esté la red y que tanta información puede reconocer en la imagen.

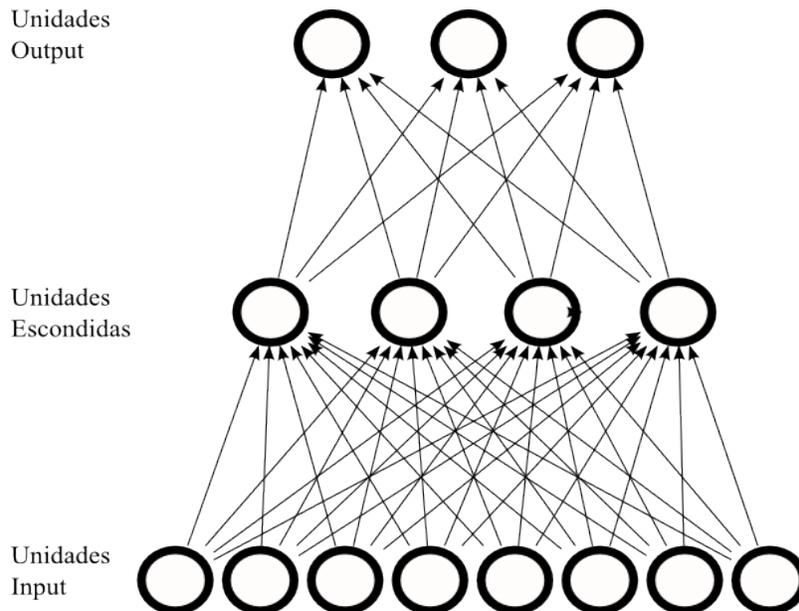


Figura 2 Un ejemplo de red neuronal de *feedforward*, donde cada unidad de *input* tiene un determinado valor, que representa la información que está entregando, los nodos escondidos con los que está conectada procesan estos datos y envían su valor de activación a la capa siguiente, donde se responde activando uno de los tres posibles valores de *output*.

Las unidades de *input* (IU) representarían las neuronas que son afectadas por un estímulo, las unidades escondidas (HU) serían las neuronas que procesan la información que éstas entregan y las de *output* (OU) son el equivalente a las neuronas que se activan para responder de alguna forma a los fenómenos del mundo. El valor que envía una IU a las HU es el homólogo a la señal que envía una neurona a otra, puede aumentar o disminuir el valor de su señal, de la misma forma que una neurona afecta la frecuencia de disparo de la señal de otra. En este sentido, lo que está ocurriendo en la red depende del patrón de activación de la misma, el cual está determinado por la función que realizan los distintos nodos dentro de la red. Por lo general esa función implica una suma de los valores aportados por cada IU a un nodo particular entre las HU, ponderado por el peso de la conexión entre el receptor y el nodo particular en la capa escondida. Este proceso también puede realizarse de manera que la

información del *output* de la red sea enviada de vuelta a las HU, de manera que ellas puedan generar un *output* contextualizado. Estos dos ejemplos de computación serán aclarados más adelante en los ejemplos de redes neurales, el primero corresponde a una red de *feedforward* y la segunda a una red con capacidad de recursividad.

II.I.I Ejemplos de redes neurales con comportamiento 'inteligente'

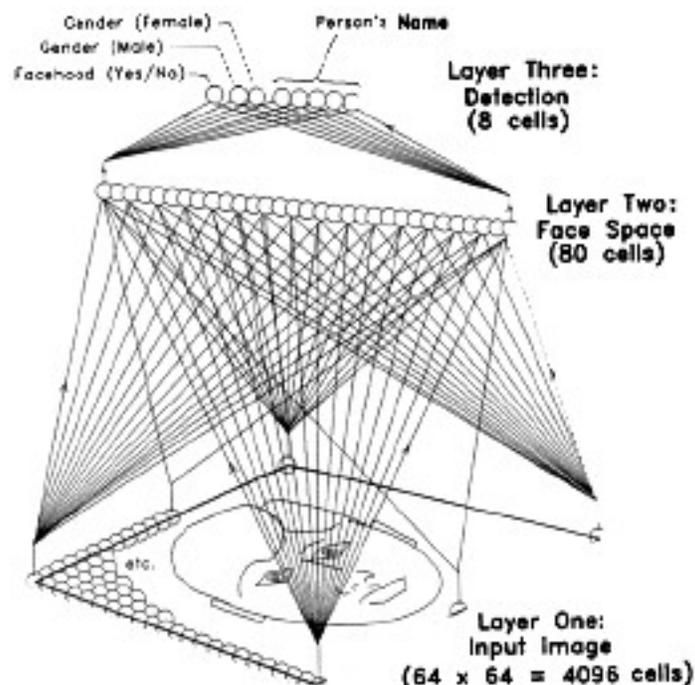


Figura 3 Ejemplo de red neuronal artificial para reconocimiento facial. Tomado de Paul Churchland, 1996.

Distintos ejemplos de red neural son presentado en Churchland 1996, en particular me parece interesante la red neural artificial capaz de reconocer rostros presentada en el capítulo 3⁸, la cual fue diseñada para esta tarea y además entregar información relevante sobre los rostros que son reconocidos, junto con diferenciar aquellos *inputs* que no lo sean. La red consiste en 3 capas: la primera capa funciona como una retina que es afectada por una imagen de 64x64 píxeles, donde cada uno puede tener 256 niveles de brillo. Cada una de las células en esta pseudo-retina proyecta su nivel de activación a todos los nodos; La segunda capa está formada por 80 nodos, cuyos pesos son

⁸ La red en cuestión no fue desarrollada por P. Churchland, sino que por Garrison Cottrell y su equipo en la Universidad de California, San Diego.

cuidadosamente calibrados de manera que discriminen en primer lugar si lo presentado es o no una cara, luego, y en caso que lo sea, si es un varón o una mujer y en última instancia, cual es el nombre de la persona que está viendo, representado por un código numérico asignado al azar. Este último paso es el *output*, que hace referencia al patrón de activación de la red frente al estímulo presentado o la representación que tiene la red de este. La red es entrenada con un set de 64 fotografías diferentes, 13 fotos de no-caras y 11 caras en distintas expresiones (Fig 4).

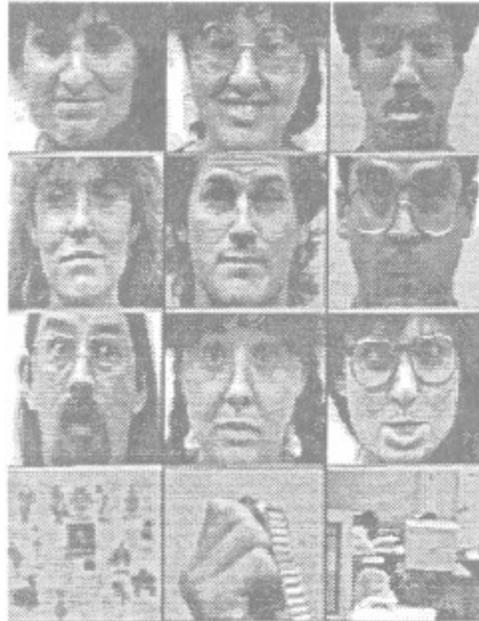


Figura 4 Ejemplos de las fotografías con que se entrenó la red. Churchland, 1996.

En un principio los pesos de las conexiones son dados al azar, no hay forma *a-priori* de calibrar los pesos de la red de manera que se asocie correctamente un *input* determinado con el *output* deseado, esta es una red que consta de 4184 células y con 328.320 conexiones, no es fácil definir el valor correcto desde el principio. La idea es que la red regule sus pesos mediante una ley de aprendizaje, en este caso se usó una regla conocida como retropropagación. Como se dijo, se comienza uniendo las células de la capa escondida con las de *output* según un valor al azar para los distintos pesos, positivos y negativos pero ninguno muy lejano a cero, luego se le presenta un *input* elegido entre las imágenes de muestra y se compara el *output* que esta primera prueba genera con el esperado. El *output* esperado en estos casos será un vector tipo $\langle 1; 0, 1; 0.5, 1, 0.5, 0, 0 \rangle$ donde los dos primeros valores indican que es verdadero que es una cara (1), que no es un varón, sino una mujer (0, 1) y que su nombre clave es

[0.5, 1, 0.5, 0, 0]. Al no estar en absoluto bien templada, la red arrojará un output distinto, digamos <0.4; 0.23, -0.5; 0.2, 0.1, 0.06, 0.3, 0.9>. Se resta el valor obtenido con el esperado y el resultado se eleva al cuadrado, finalmente se determina el promedio de esos valores y al resultado obtenido se le denomina error promedio cuadrado (MSE), es decir:

<	1,	0,	1,	0.5,	1,	0.5,	0,	0	>	<i>Output</i> esperado
-<	<u>0.4,</u>	<u>0.23,</u>	<u>0.5,</u>	<u>0.2,</u>	<u>0.1,</u>	<u>0.06</u>	<u>0.3,</u>	<u>0.9</u>	>	<i>Output</i> obtenido
	0.6,	-0.23,	0.5,	0.3,	0.9,	0.44,	-0.3,	-0.9		Vector de error
	0.36,	0.05,	0.25,	0.09,	0.81,	0.19,	0.09,	0.81		Error cuadrado
				0,3313						Error Promedio Cuadrado

El MSE representa qué tan equivocada está la respuesta de la red frente al *input*, lo que se debe buscar es una forma de que este disminuya lo más que se pueda. Para esto, se mantienen los pesos de toda la red, salvo el de una conexión, el cual se varía un poco de manera que sea posible determinar lo relevante de esa conexión para la asociación de ese *input* con el *output* deseado. Luego se vuelve a comparar el *output* obtenido con el ideal y en caso de que el MSE sea menor, se pasa al peso inmediatamente siguiente. Se continúa así con cada uno de los pesos y se realiza este procedimiento con cada una de las imágenes del conjunto de entrenamiento, hasta que el valor del MSE es lo más bajo posible, llegado este punto se tienen calibrados los pesos de la red y el entrenamiento puede considerarse terminado.

Es claro que la red debe tener un buen resultado con las imágenes que ha sido entrenada, de hecho consigue un 100% de aciertos con el conjunto de entrenamiento, los pesos están ajustados de manera que lo haga correctamente, pero esto no implica que reconozca a las personas en la imagen, cabe la posibilidad que haya 'memorizado' de alguna forma el conjunto con el que fue entrenada. Los resultados interesantes aparecen al presentar fotos nuevas de esas caras, imágenes fuera del entrenamiento pero de las mismas personas a quienes se le entrenó para reconocer, en este caso obtiene un 98% de eficacia, confunde el nombre y género de una mujer.

Más interesante aun es lo que ocurre cuando se le presentan imágenes completamente desconocidas para la red. En estos casos la red acierta en un 100% al determinar si el *input* es o no una cara y tiene un 81% de aciertos al determinar el género de la cara que se le presenta, por algún motivo confunde a

mujeres con varones pero no viceversa. Incluso cuando se le presentan las caras conocidas pero con 1/5 del rostro oscurecido mantiene similar porcentaje de aciertos, salvo en el conjunto de imágenes en que el quinto censurado era la frente de los sujetos, donde el rendimiento disminuyó a 71%, mostrando que la posición del cabello es relevante en su proceso de clasificación.

El secreto para conseguir tan buen desempeño tiene que ver con la manera en que se codifica la información en la segunda capa. Esas 80 células constituyen un espacio de representaciones faciales de 80 dimensiones, donde cada una de los rostros con los que fue entrenada ocupa un lugar único, especificado por un patrón de activación de todos los nodos dentro de la capa. Ahora el problema tiene que ver con cómo es que se organiza este espacio, ¿cómo es que reconocen que una determinada imagen es tal o cual persona o que tiene tal o cual género? No parece implausible postular que cada una de las células se dedica a reconocer algún patrón específico dentro del estímulo recibido, el tamaño de la nariz, la distancia entre los ojos o alguna otra característica particular. Una de las ventajas de que estos modelos sean implementados en computadoras, es que permite determinar cual es el patrón de activación que genera su máxima respuesta según su relación con la capa que le precede y con esta información reconstruir el estímulo preferido de cada una de estas células.

El resultado de esta reconstrucción, en la figura 5 se pueden apreciar 6 de ellas, está lejos de la primera intuición, al mirar estas reconstrucciones se puede apreciar que cada nodo se fija en toda la superficie de la capa de *input*, de manera que cada una representa una estructura similar a una cara completa, en vez de una propiedad aislada de ella. Pero esto no implica que cada uno de estos estímulos preferidos corresponda a una cara particular del conjunto de entrenamiento, más bien parecen una variedad de rasgos *holísticos*. Al ser afectados por una cara, cada uno de los rasgos *holísticos* se activará de una determina manera y responderán emitiendo un patrón de activación que será único para ella. Imágenes distintas de una misma cara producirán esencialmente el mismo patrón de activación, permitiendo que la capa de *output* responda de manera correcta.



Figura 5 6 estímulos preferidos, *holones*. Se puede apreciar que son imágenes que reconocen toda la superficie del *input*. Churchland, 1996.

Este es un ejemplo de representación distribuida, la información no es almacenada en estructuras atómicas, sino que está dispersa en todos los nodos de la red. Los distintos nodos dentro de la capa escondida son las distintas dimensiones del espacio de representación de la red, podríamos entender un *holón* como el punto máximo de cada una de ellas, y la activación que produce el estímulo que es recibido implica un punto entre la no-activación y el *holón*.

Una nueva propiedad emerge de esta manera de codificar la información, una estructura jerarquizada de categorías reconocibles. Si dibujamos solo 3 de las 80 dimensiones del espacio perceptivo de las células (Fig. 6), no se pierde información con esto y ayuda a tener una imagen del fenómeno, es posible apreciar que debido al entrenamiento que ha recibido, la red separa las representaciones que puede formar en distintas particiones, la primera división, la que se produce más cerca del origen, es donde se codifican los objetos que no son caras, al no poder reconocerlos el patrón de actividad que generan es bajo. La partición que codifica los rostros está dividida a su vez en dos partes más, una región que aglomera las caras masculinas y la otra las femeninas, en ambas puede encontrarse un “centro de gravedad” que representa el prototipo de rostro de esa subregión y a medio camino de ambas representaciones arquetípicas, se encuentra una cara de género ambiguo, con rasgos que podrían pertenecer a cualquiera de las dos subregiones. Esta diferenciación en distintas regiones muestran cómo es posible el surgimiento de conceptos y categorías, lo que aún cuando ocurre de manera rudimentaria, entrega una luz sobre el modo en que se podrían formar estas en nuestras propias redes neuronales.

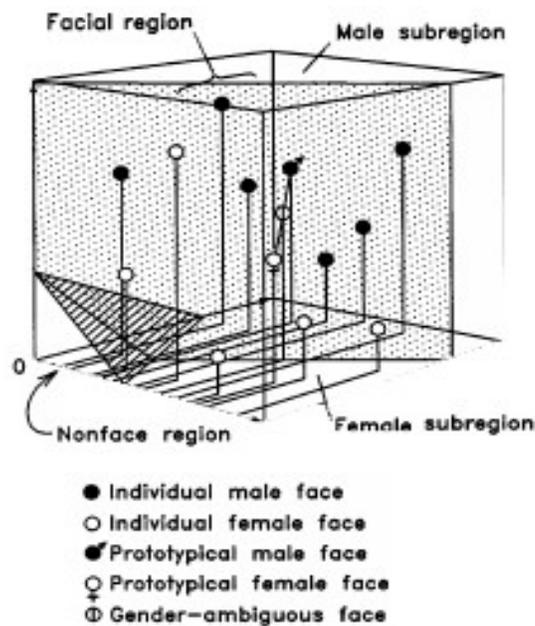


Figura 6 Visualización del espacio representativo de la red neural, con sólo 3 de las 80 dimensiones se puede apreciar cómo los patrones de activación se organizan de manera que dan origen a categorías claramente diferenciables, aun cuando esto no fue parte del entrenamiento dado a la red. Churchland, 1996.

Los rostros prototipo de la red dependerán del conjunto de entrenamiento, en la presentación de estos modelos se citan otros experimentos⁹ en los que a redes similares se les entrena con conjuntos de imágenes de rostros mayoritariamente orientales y algunos caucásicos, masculinos y femeninos en ambos casos, la red responde de manera similar a la ya descrita. Es decir, organiza las imágenes en categorías claras, con regiones para cada distinto grupo de representaciones con similitudes, pero pierde efectividad al discriminar los géneros de imágenes nuevas cuando estas son de rostros caucásicos, les parecen casi lo mismo. Ocurre algo análogo al entrenarla con una mayoría de rostros africanos y unos cuantos orientales, se desenvuelve muy bien con los rostros novedosos africanos, pero mal con los orientales.

En psicología se llama a este fenómeno “efecto de familiaridad”, nos resulta más fácil reconocer rostros con rasgos a los que estamos acostumbrados, pero al enfrentarnos a situaciones novedosas nos cuesta diferenciar a un individuo u otro¹⁰. Si bien no es especificado en el texto, creo que no atenta contra ninguna ley del pensamiento el postular que esta red entrenada con rostros tan variados, donde no solo cambia el género, sino también muchos otros rasgos característicos. Esta también ordenará los

9 Realizados por Alice O'Toole y su grupo en la universidad de Texas, Dallas.

10 Un ejemplo de esto es el chiste que dice que todos los orientales se parecen a Jackie Chan.

rostros que se representa en categorías jerarquizadas, pero en estos casos se podrían apreciar aun más divisiones dentro de las subcategorías. Donde, además de separar a varones y mujeres, habría una subdivisión para los rostros caucásicos, orientales y africanos, cada uno de estos subconjuntos con sus propios “centros de gravedad” y organizaciones.

La ventaja de esta manera de codificar la información no radica en su eficiencia solamente, además permite que la red pueda mantener su funcionamiento aun cuando algunas células sean dañadas o algunas conexiones perdidas. En la medida que cada pixel del *input* tiene un efecto, por mínimo que sea, en cada nodo de la segunda capa y cada nodo tiene algo de información relevante sobre la situación de todo el *input*, la pérdida de células dispersas o la desconexión de algunas relaciones no implican que la red deje de funcionar, quizás sea menos prolija, pero sus resultados seguirán siendo confiables.

Parece claro que este tipo de red neural se comparta de manera similar a la que lo haríamos nosotros, el problema está en que la regla con que se regulan los pesos de las conexiones es poco realista en términos biológicos, pues si bien en nuestro cerebro las sinapsis se van reordenando y varían con regularidad, no es claro que lo hagan según el modelo propuesto por la retropropagación. Es decir, no vamos aprendiendo cambiando las conexiones entre una neurona y otra para relacionar un *output* con un *input*, si no que esto sucede de manera natural y sin mediar MSE alguno. Además de esto, está el problema de la recursividad, las capas de los niveles superiores no son capaces de afectar a las de más abajo, generando atención a una determinada zona del *input* o memoria a corto plazo reteniendo la información relevante por periodos cortos de tiempo, cómo sí ocurre en el cerebro. Antes de continuar con el siguiente ejemplo, intentaré aclarar un poco el problema de la recursividad, pues será relevante para comprender la segunda red y para resolver los problemas que se afirma tiene el modelamiento conexionista.

En la figura 7 puede verse un ejemplo de red recurrente, donde algunos de los nodos de la capa de *output* están conectados con los nodos en la capa escondida, de manera que funcionan como contextualizadores. Al recibir un primer *input*, la primera capa comunica esta información a la capa escondida, esta a su vez entrega la información a la capa de *output*, la que envía parte de esta información de vuelta a la capa escondida y al momento que la capa de *input* recibe un segundo estímulo, la capa escondida puede relacionar este nuevo *input* con la información que ha retenido gracias al envío de la información de la capa de *output* en el momento anterior. Esto da a la red la

posibilidad de trabajar no solo en el ahora, sino que le da un pequeño espacio de memoria hacia el pasado, contextualizando la nueva información con la que ya ha recibido.

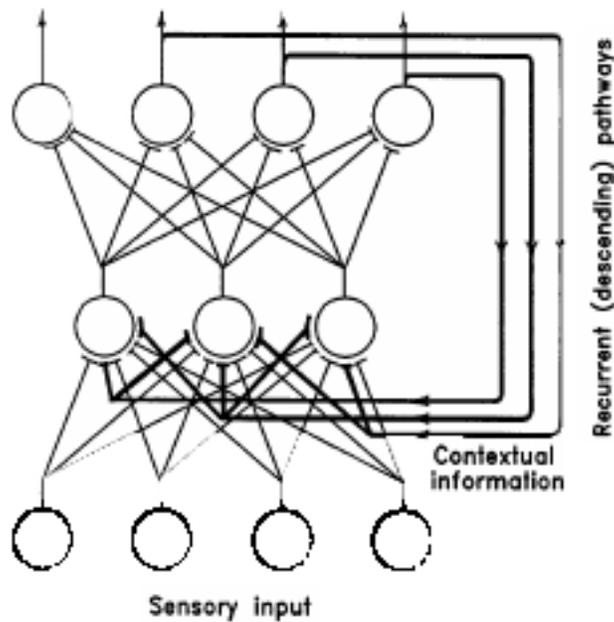


Figura 7 La información recibida por la capa oculta, luego del primer estímulo, es analizada junto a la información que la red retiene de su anterior respuesta. De manera que la respuesta a esta segunda instancia depende de la respuesta anterior y por lo mismo del *input* recibido anteriormente. Churchland, 1996.

Dentro de las ventajas de esto está la posibilidad de reconocer relaciones causales, pues al ser capaz de relacionar distintos *inputs* con sus *outputs* respectivos y producir *outputs* con la ayuda de la contextualización del *output* anterior, se produce una relación temporal entre los nuevos *inputs* y los *outputs* anteriores, de manera que estos pueden ser considerados como un mismo objeto cambiando de posición en el espacio sensible, como en el ejemplo presentado en la figura 8. La información en estas redes es almacenada en nuevos tipos de categorías, no solo son relevantes las relaciones entre las representaciones que se reconocen, sino que se comienzan a reconocer direcciones dentro de los espacios de activación, representación, de manera que la red es capaz de asociar distintas representaciones según patrones temporales y estas son capaces de predecir un futuro estado del *input*, o cómo debería continuar su trayectoria, de manera no arbitraria.

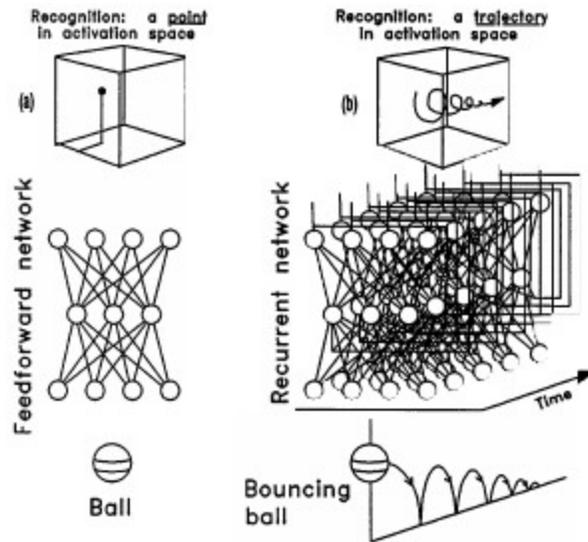


Figura 8 La red (a) es capaz de reconocer una pelota, esta afecta a la red de una manera específica y, por lo mismo, tiene un lugar determinado dentro del espacio de activación. Esa misma pelota puede ser reconocida por la red al moverse en el espacio y el tiempo, si la red es capaz de comunicar su *output* a la capa anterior, de manera que el *input* nuevo es contextualizado, con esto el movimiento de la pelota no es reconocido como distintas pelotas en distintos puntos del espacio, sino como una misma pelota rebotando. Churchland, 1996.

III Problemas asociados al conexionismo: sistematicidad y productividad de los procesos mentales.

Los modelos conexionistas no están libres de críticas, desde sus orígenes se han presentado distintos argumentos contra ellos. Los principales argumentos contra esta forma de interpretar las representaciones mentales y los procesos cognitivos tiene su origen en la manera de describir estos fenómenos con la que compete. La idea que defiende este otro grupo es en muchos sentidos similar a la sostenida por los conexionistas, defienden la existencia de los fenómenos mentales y la naturaleza compleja de ellos, además se adhieren al materialismo respecto a la mente, pero difieren en la manera de explicar la naturaleza y funcionamiento de los fenómenos mentales. La otra interpretación de la que hablo es la llamada clasicista, que sostiene que las RM tienen una sintaxis y semántica combinatoria, según esto las RM serían estructuras de símbolos, distinguibles entre representaciones complejas o simples, donde las primeras estarían formadas por las segunda y donde el contenido semántico de una representación compleja esta en función al contenido semántico de sus partes sintácticas. Estas características implican que los procesos mentales, al ser relaciones entre RM, son estructuralmente sensibles, es decir, la manera en que se relaciona un *input* con un *output* está definido por las propiedades estructurales de la representación mental que implica el primero. Al estar comprometidos con el realismo, supone que la estructura combinatoria de una RM es equivalente a relaciones estructurales entre propiedades físicas del cerebro, donde los símbolos en la RM corresponden a estructuras físicas del cerebro¹¹. Y los modelos que implementan esta descripción tienen estructuras que respetan las arquitecturas de Turing y Von Neumann. Estas propiedades de las RM y de los procesos mentales está ligada con la hipótesis del lenguaje del pensamiento, que implica que los procesos mentales son operaciones de transformación y combinación de estructuras simbólicas mediante reglas, por esto se dice que la teoría clasicista se sostiene en una comprensión lógico/sintáctica de las RM.

Jerry Fodor y Zenon Pylyshyn han desarrollado estas hipótesis de maneras bastante plausibles, logrando descripciones de procesos cognitivos que parecen satisfactorias, además han presentado argumentos que han sido relevantes a la hora de considerar las teorías conexionistas y su valor para la psicología o la filosofía de la mente. Las características que Fodor y Pylyshyn¹² indican como las principales fallas en las descripciones conexionistas son la ausencia de 4 propiedades, a saber: La

11 Respecto a la plausibilidad de esta interpretación del cerebro no he encontrado información al respecto, pero el resto de las publicaciones y descripciones de su funcionamiento no hacen claro que esta sea una interpretación plausible.

12 Fodor y Pylyshyn (1988)

productividad del pensamiento; la sistematicidad de las RM; la composicionalidad de las RM; y por último, la sistematicidad de la inferencia. Las cuales son propiedades básicas de nuestros procesos cognitivos y sus estructuras constituyentes.

La productividad del pensamiento es apreciable en la posibilidad de crear una cantidad de proposiciones aparentemente infinita, según parece nuestro sistema representacional no tiene límites en cantidad de representaciones posibles. Con un número finito de elementos somos capaces de generar, probablemente, infinitas expresiones, esto sólo puede conseguirse si las expresiones son compuestas y los elementos dentro de ellas son transportables, lo que implicaría que las representaciones lingüísticas (RL) y mentales se constituyen como sistemas de interacción de símbolos.

La sistematicidad de las RM se puede ver en nuestras capacidad lingüísticas, que también son RM, y refiere a la habilidad de producir/comprender algunas oraciones está intrínsecamente relacionado con la habilidad de producir/comprender otras. Por ejemplo: alguien que es capaz de entender o producir una oración como “Juan quiere a María”, seguramente es capaz de comprender/producir la oración “María quiere a Juan”. Esta propiedad se relacionaría con un conocimiento de la sintaxis de un idioma, no es una propiedad de las palabras que lo componen, no parece haber una propiedad en Juan o en María que hagan posible pensar que son intercambiables en la oración del ejemplo anterior.

La composicionalidad de las RM tiene que ver con que, pese a no tener claras las reglas de cómo lo hacemos, somos capaz de organizar oraciones con sentido sin mayor problema. Lo que implicaría necesariamente una sintaxis y semántica para este tipo de representaciones, como las RM también funcionan de este modo, también la tendrían. Por ejemplo, si no hubiese una semántica y sintaxis necesaria y expresable en términos de relaciones entre símbolos y funciones, no habría forma de formar la oración “Juan iba caminando por la calle” con las palabras “Juan”, “ir”, “caminar”, “por”, “la” y “calle”.

La sistematicidad inferencial, por su parte, refiere a las relaciones nomológicas que hay entre las representaciones lingüísticas (RL) y se relacionan con las RM pues parece necesario que si una hay un pensamiento que implica una RM del tipo AyB, es necesario que haya un posible pensamiento que

implica la RM A y otro pensamiento posible que implica B y esto debe cumplirse para cualquier otra relación lógica en cualquier individuo que maneje el lenguaje. Y estas deben corresponder en términos de su rol inferencial dentro una RM compuesta, no es posible que alguien pueda pensar PyQyR e inferir Q, pero no pueda hacer lo mismo desde PyQ, otro sujeto con las mismas características en el modo de procesar, almacenar y comunicar información tampoco debería ser capaz de esta falla.

Es claro que esta clase de interpretación acerca mucho la discusión sobre las RM a un modelo del pensamiento donde este funciona a la manera en que asumimos que funciona el lenguaje, de hecho, quienes proponen estas críticas son defensores de la hipótesis del LDP. Ya que las funciones mentales proceden como un lenguaje, donde las RM más simples son símbolos y las más complejas relaciones de estos, debe haber un LDP y esta es la forma de modelar la mente. El contenido semántico de una RM está determinado por sus relaciones con otras y, gracias a una gramática natural, las distintas instanciaciones de objetos mentales son de alguna forma activadas por relaciones con otras instanciaciones y esa relación tiene que ver con la manera en que funciona el mecanismo que instancia el objeto. Estas instanciaciones funcionan como símbolos, que se presentan como activaciones en el cerebro y la forma en que esas activaciones se relacionan depende de las reglas con las que funciona el cerebro.

El problema es que esas mismas características podrían ser defendidas desde el conexionismo, pues también pretende -y debe- dar cuenta del lenguaje natural, lo que implica que necesita que su teoría explique las propiedades de este y su relación con las RM caracterizadas en él. Esto es: cómo se relacionan las RL con las RM. En lo posible, dar un modelo que cumpla estas reglas y que funcione a la manera en que se comporta el cerebro, que muestre las propiedades semánticas y sintácticas del lenguaje, además de ser capaz de explicar esa relación en términos de procesos del cerebro, junto con permitir que este modelo sea implementable en una máquina, de manera que esta sea capaz de formar, al menos en principio, infinitas frases e infinitamente complejas. El conexionismo tiene que poder explicar la forma en que estas frases se relacionan entre sí y cómo obtienen contenido. Además de permitir contar una historia de cómo es que se puede dar esto, cómo y porqué es posible esto.

Primero es necesario comprender que las RM son parte de todo sujeto con vida mental y si asumimos que función mental y función cerebral son distintos niveles de explicación para un

fenómenos similar, los sujetos con SNC similar al nuestro tendrán RM, de hecho hay evidencia de mamíferos capaces de lenguaje, se comunican entre ellos mediante el intercambio de señales y estas son regulares dentro de una comunidad. Pero principalmente, si está sujeto a una historia de desarrollo, es claro que el análisis de las propiedades de la dinámica del cerebro, junto con el análisis de la dinámica del lenguaje, nos dirán más sobre la dinámica de las RM y las RL. Parece claro que las RL son un subconjunto de las RM, según esto y por reglas propias de los conjuntos, el denso conjunto de las RL posibles, debido a sus propiedades antes mencionadas, si bien tiene una cardinalidad infinita, esta es infinitas veces menor al conjunto de RM posibles¹³. Nosotros nos representamos al mundo no sólo mediante el lenguaje, también lo escuchamos, lo olemos y lo miramos, el cuento del mundo que nos contamos está constituido de mucho más que nuestras maneras de describirlo discursivamente¹⁴. Claramente el lenguaje natural permite dar cuenta de algunas cosas que ocurren en la relación entre los procesos mentales y permite dar cuenta de algunas propiedades de ellos el conocer la formalización de las regularidades en este tipo de representaciones. Pero el hecho que su estructura y reglas de conformación sea parte de las reglas y estructuras que necesariamente tiene que explicar una teoría de la representación con miras a un modelo mental, no implica que sea necesario dar la explicación de todas las RM en términos de ellas. Por otra parte, si conocemos las leyes que gobiernan las RM, o pretendemos tenerlas, debemos ser capaces de que producir un modelo que pueda aprender a generar RL y que estas tengan las propiedades que ya fueron mencionadas. Como el conexionismo pretende tener alguna idea de estas leyes, debe ser capaz de implementar un modelo donde la semántica y la sintaxis de las RL se vería subsumida en la sintaxis de nuestro cerebro en términos de cómo opera la transformación, almacenamiento e implicación de patrones de activación de las conexiones de los elementos que lo componen, una neurosemántica y una neurosintaxis.

Mostrar que es posible la productividad y sistematicidad se haría presentando un modelo conectivista que, en base a un conjunto de palabras aprendidas en contextos independientes, fuese capaz de generar predicciones u oraciones con estructura sintácticamente correctas, con categorías gramaticales claras y ser capaz de relacionar los elementos aprendidos como instanciaciones de representaciones similares, generalizando reglas de manera de producir estructuras nuevas cuya

13 Del mismo modo los números naturales y los racionales son subconjuntos de cardinalidad infinita de los números reales. Aun cuando los números racionales son infinitamente más que los naturales, entre 1 y 2 hay infinitos números racionales, pero ningún natural.

14 Un gran personaje de la historia hizo decir a otro personaje algo como: “Hay más cosas entre el cielo y la tierra que lo que sueña tu filosofía”. Frase que me parece refleja esta idea de manera bastante atractiva.

organización interna respete la gramática del lenguaje con el que fue entrenada. Demostrar que es posible la sistematicidad y la productividad en un modelo conexionista solucionaría los problemas presentados, pues estas dos propiedades son la base del resto de las críticas, es en base a ellas que se produce la sistematicidad inferencial y la composicionalidad.

La idea es encontrar una respuesta desde el conexionismo, por lo que se asume que un modelo desarrollado a la manera clasicista respetaría estas constricciones, pero este perdería su relación y compromisos con el naturalismo, al no modelar la computación mental en base a la forma en que se trabaja la información en el cerebro. La solución estaría en mostrar que es posible generar productividad y sistematicidad en un modelo conexionista, pues si somos capaces de esto, tenemos una gramática y semántica, con lo que se daría cuenta de las otras propiedades presentadas, al menos en principio. Pues las cuatro propiedades de las RM/L presentadas por Fodor y Pylychyn se sustentan en que es gracias a esa semántica y gramática que se producen y relacionan esos fenómenos, por lo que en última instancia hay que asumir que las RM se comportan como frases en un lenguaje.

Ya se mencionaron ejemplos de redes recursivas, donde el *output* de un momento afecta el *output* que puede seguir en otro, lo que genera algo similar a memoria de corto plazo, de modo que la red trabaja no solo reconociendo patrones, sino que a la vez puede presentar predicciones sobre *inputs* y completarlos en caso de que estos no lo estén. El modelo de red conexionista recursiva desarrollado por Jeff Elman¹⁵ sirve para comprobar si es posible que estas aprendan reglas gramaticales, sin que estas sean explícitamente introducidas al modelo, y que construyan oraciones de mayor complejidad de manera gramaticalmente correcta, aun cuando no se le haya enseñado explícitamente cómo debe relacionar distintas oraciones. Esto se podría apreciar en una red que fuese capaz de predecir predicados u otras partes constituyentes de una oración sin que se le expliciten y de ordenar sus representaciones en categorías gramaticales relevantes, mientras lo segundo de mostraría si una red es capaz de predecir como continuar oraciones de mayor complejidad, en comparación con las presentes en las oraciones del conjunto de entrenamiento.

El modelo que se presentó para atacar el primer problema, de las categorías gramaticales, era algo más menos similar al mostrado en la figura 8(a). Este fue entrenado con 29 palabras, entre verbos

¹⁵ Este tipo de red fue desarrollada a mediados de los 80 por el profesor Elman, son conocidas como redes elman o TRACE.

y sustantivos, y un total de 10.000 oraciones de 2 o 3 palabras. El *input* que la red recibía era de una palabra por vez y su tarea era predecir la que debía seguir en relación a la anterior. Se asume que no hay una sola predicción correcta, lo importante es que no prediga un verbo al momento de dar un objeto directo o problemas gramaticales similares. El segundo problema, de la relación entre gramática y productividad, fue trabajado con la red idealizada en la figura 8(b), que fue entrenada con 8 sustantivos, 12 verbos y según un conjunto de reglas gramaticales, se formularon 10.000 oraciones con sentido, cuya complejidad era dada por la cantidad de oraciones relativas de las que estaba compuesta, la idea es que fuese capaz de predecir todas las palabras que eran gramaticalmente admisibles como continuación para un determinado *input*, es decir que conjugara verbos, que organizara una oración relativa y procesos similares.

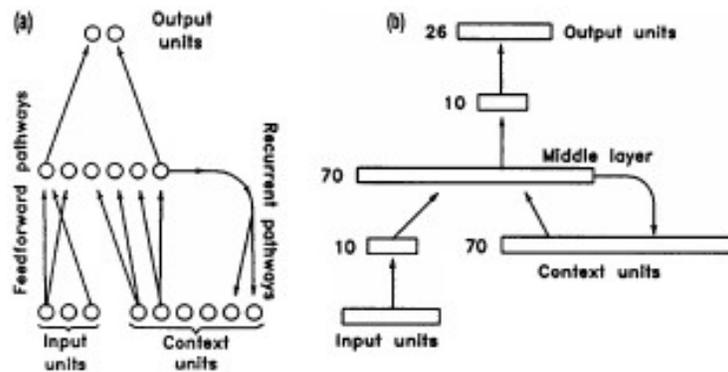


Figura 8 (a) Red recursiva simple, donde a cada *input* la capa de en medio relaciona células contextuales, lo que permite que puede poner el *input* en relación con las representaciones que tuvo en un momento anterior. Luego, puede evaluar sus nuevos *inputs* en vista a las secuencias de palabras dentro de las que ha aparecido. (b) Una red un poco más compleja que pretende poder discriminar gramaticalmente entre oraciones con clausulas relativas, relacionar en categorías oraciones complejas con clausulas relativas. Churchland, 1996.

El comportamiento de la red de la figura 8(a) fue bastante bueno, las predicciones no siempre eran del todo elegantes, pero no predecía un verbo como continuación de otro, era capaz de entregar complementos directos cuando era necesario, en general, no erraba en la categoría gramatical de la palabra que debía seguir a la oración incompleta del *input*. El resultado del análisis de la capa del medio en la red en la figura 8(a), mostró que las representaciones que esta generaba estaban organizadas según el diagrama de la figura 9, se le agregaron los nombres a las subdivisiones luego de obtenido el diagrama, donde es claro que la red fue capaz de discriminar las distintas categorías

gramaticales relevantes que le fueron presentadas y además de agrupar los elementos dentro de estas categorías según sus relaciones internas, separando entre grupos según categorías semánticas.

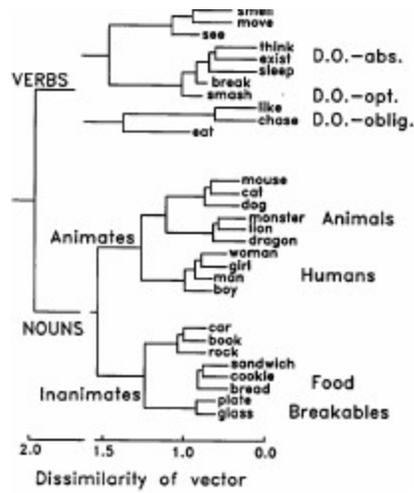


Figura 9 Organización de las representaciones que produce la red de 8(a), con una categorización según características gramaticales y semánticas. Churchland, 1996.

El modelo que se usó para tratar el segundo problema mencionado también respondió de manera sorprendentemente parecida a lo que se esperaba. Produciendo correctas conjugaciones y relaciones entre verbo-sujeto, si bien fue entrenado en inglés donde los verbos regulares solo cambian en la conjugación de la 3ra persona singular esto no deja de ser un buen comienzo, además de ser capaz de mantener las estructuras entre las oraciones relativas y predecir resultados que podían tener clausulas relativas, bien organizadas, a su vez. El análisis de los datos de la capa del medio presentado en la figura 10 muestra que las oraciones con componentes similares tienen maneras de moverse en el espacio de representación similares, se mueven de manera similar, pero su sincronía se pierde cuando varía alguna parte de la oración.

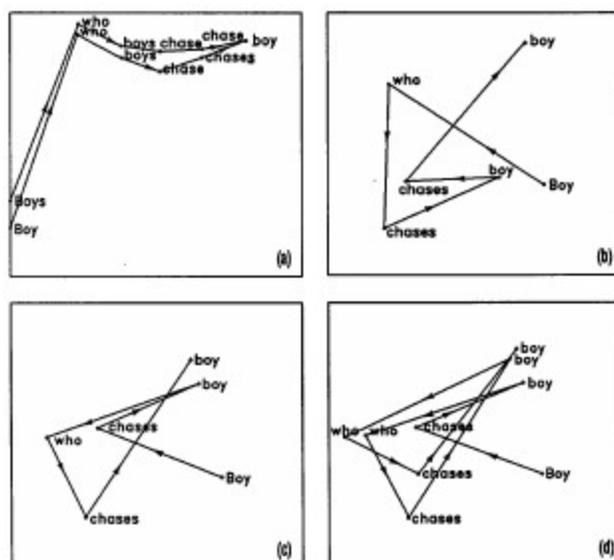


Figura 10 Se presentan distintas maneras de relacionar los elementos en una oración, en la tarea de completa las oraciones, el movimiento de una palabra hacia la siguiente muestra como la ordenación de la información dentro de la red no es arbitraria, las relaciones de complementos directos, verbos y oración relativa son representados constantemente con movimientos similares dentro del espacio lingüístico de la red.

En el cuadro (a), por ejemplo, se muestra que oraciones similares tienen una similar dirección en el espacio, además de un origen cercano semántica y sintácticamente, continúan esa simetría hasta variar en la activación del verbo relativo al sujeto en que se diferencian y luego concluyen en un objeto directo similar. Esto es una muestra de que el modelo conexionista es capaz de dar resultados esperados y que la manera en que los produce es, metafóricamente, similar a la nuestra. Relacionando semánticamente los elementos en el lenguaje y generando reglas sobre formación de oraciones mediante generalizaciones sobre casos particulares y aparentemente no relacionados.

La evidencia que entregan estas redes sobre las posibilidades de las redes conexionistas es clara, son capaces de organizar sus representaciones en categorías y aprender reglas sobre cómo organizar de manera coherente estructuras complejas con ellas. Según parece, los problemas asociados con ellas tiene más que ver con la mala interpretación de lo que es una red conexionista que con una falencia de las mismas. De hecho, la línea argumental podría haber comenzado desde ese punto, pero lo relevante no es mostrar el error, si no cual es la solución.

El problema que queda por resolver es el origen de las RL, si bien no es un problema que se

exija responder, los modelos clasistas y la interpretación serial de la mente entregan casi con su descripción una manera de explicar cómo se producen estas, la idea de una mente/cerebro 'hechos para' organizar de forma racional los símbolos que produce, en conjunto con el background chomskyano, la modularidad de la mente y una gramática natural, hacen casi automático un lenguaje proposicional. Mientras tanto, los conexionistas pueden decir mucho sobre la organización de las áreas del cerebro y patrones de activación, pero: ¿Cómo es que puede un conjunto de sonidos relacionarse con el mundo? Este problema no es solucionable mediante un ejemplo de red o algo similar, mas el *background* teórico sobre el que se sustentan los modelos conexionistas, como se ha comentado con anterioridad, pretende respetar las explicaciones de la ciencia, en particular de la neurociencia, y desde ellas buscar hipótesis que puedan responder a los problemas presentados por la filosofía de la mente. Este caso no es una excepción, los filósofos conexionistas y naturalistas deben buscar una respuesta para la relación lenguaje-mundo en alguna propiedad del cerebro. Desde hace pocos años se ha ido desarrollando una propuesta, que si bien contrasta con la gramática natural chomskiana y la idea de la modularidad de la mente, se sostiene en la capacidad del cerebro de relacionar distintos tipos de RM de maneras no arbitrarias, todo esto derivado de los conocimientos que ha entregado el estudio de sujetos sinestésicos.¹⁶

La sinestesia es una rara condición, en realidad menos rara de lo que uno pensaría en términos estadísticos, donde un estímulo sensorial de un tipo es asociado a la vez con otro, por ejemplo, un sinestésico experimentará un color junto con un número, 5 es rojo, o un olor junto con un tono, el do es azul. Dentro de la imaginaria del sinestésico, el estímulo que gatilla la sinestesia es siempre asociado con el mismo estímulo anómalo, esto es, si el número 5 es visto de un color, cada vez que se ve 5 se asocia con el mismo color, esta condición ha sido conocida por más de un siglo, pero por lo general era poco estudiada por asumirse que era una broma, producto de la locura o un dato anecdótico en la historia de la ciencia¹⁷. Gracias a los datos obtenidos al estudiar estos sujetos, se ha mostrado que el cerebro posee mecanismos anteriores al lenguaje, anteriores en términos de historia evolutiva de los seres vivos, que muestran una conexión no arbitraria entre objetos y sonidos, como en el caso de buba y kiki mostrados en la figura 11. Estos datos han evidenciado, también, una relación de sinestesia motor-fonética entre los movimientos de los labios y la lengua, similar a lo que se puede ver al bailar

16 Los argumentos que se presentan pueden ser profundizados en Ramachandran y Hubbard (2001) y en Ramachandran (2003)

17 Una explicación más completa de este fenómeno y los mecanismos de percepción ver Ramachandran y Hubbard 2001 o 2003, además de una descripción más completa de lo explicado de manera tan sucinta aquí.

donde los movimientos del cuerpo se relacionan con los ritmos de la música, por algo somos capaces de decir cuando alguien baila mal, cuando sus movimientos no responden correctamente al estímulo auditivo. En conjunto con una correspondencia entre apariencia y vocalización, que es posible apreciar en la relación entre las letras dentro de palabras y sus significados, por ejemplo la relación de las palabras diminutivas o relativas a cosas pequeñas y la letra i en distintos idiomas, v.g. 'petite', 'chiquito' o 'little', estas 3 características, además de las evidencias que han mostrado los estudios de resonancias magnéticas funcionales, donde se hacen notorias la activación cruzada de mapas motores encargados de la gesticulación y vocalización, son suficientes para permitir la emergencia de un proto-lenguaje vocalizado. Este luego podría desarrollarse y complejizarse según las presiones de selección que implica el hecho de ser usado como método de comunicación entre individuos en una comunidad.

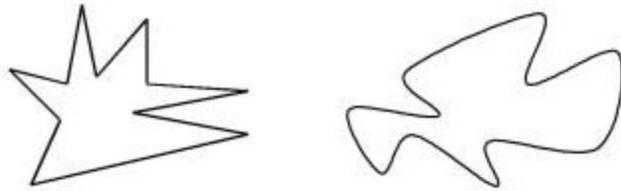


Figura 11 Estos dos símbolos son letras de un alfabeto extraterrestre, una es buba y la otra es kiki. ¿Cual es kiki y cual es buba? Si bien puede ser distinto, lo normal es que se diga que la imagen puntiaguda es kiki y la redondeada es buba, las puntas en la primera imagen se relacionan con la inflexión de la lengua al pronunciar 'kiki' y la redondez de la segunda se relaciona con lo ondulado de 'buba'. Esto muestra como relacionamos distintos tipos de representaciones entre ellas y en muchos casos estas relaciones no son del todo azarosas. Esta clase de relaciones podría dar una idea de como producimos metáforas y ser parte de lo que nos permitió en algún momento comenzar a comunicarnos con un lenguaje tan complejo como el que hemos desarrollado.

Conclusión, un experimento mental

La idea de un modelo conexionista es reproducir lo más fielmente las implicaciones de los modelos naturalistas de interpretación de las ciencias. La red que reconoce rostros podría hacer otra cosa con su *output*, además de imprimir un patrón de números, podría perfectamente enviar esa información a una red que pronunciase el nombre, que tendría que ser traducido por una red que reconociera ese patrón de números como un nombre, en particular el de esa persona que esta impresa

en la relación de activación de la red que veía la cara. De hecho, supongamos que alguien es capaz de crear una red de redes. Donde una es capaz de reconocer y producir oraciones gramaticalmente correctas, otra es capaz de reconocer objetos en el mundo y asignarles un nombre, otra es capaz de reconocer patrones complejos y persistentes, relaciones causales, otra es capaz de mover un montón de apéndices, otra es encargada de buscar relaciones entre los fenómenos presentado a la red y puede construir predicciones sobre futuros estados de la misma, otra controla la energía que recibe cada red en función a la actividad que está realizando, pues un bicho así para mantenerse en funcionamiento, necesita energía.

Esta red es mantenida en funcionamiento por un ingeniero eléctrico que la hizo en sus ratos libres, tenía suficientes trastos para jugar en casa y buscando en internet encontró distintas páginas con la información, liberada bajo licencia de *creative commons*¹⁸, sobre cómo construir y ordenar los distintos elementos en estas redes y el código de los programas que debía correr en ellas estaba bajo licencia *GPL*¹⁹. El resto fue cautín, copiar-pegar y compilar.

Asumiendo que estas redes se comportan cómo lo hacen las ya expuestas:

¿Se puede explicar el comportamiento de una red con esas características como intencional?

Continuando con la suposición, digamos que un día le presento distintos recortes de imágenes de cuadros pintados por distintos autores y comienza a agrupar los que fueron pintados por Mondrian, ¿Le gusta Mondrian?

Según lo dicho anteriormente y sirviéndome de la idea presentada por Dennett²⁰ sobre la intencionalidad como un método de análisis del comportamiento, como un compromiso metodológico y no necesariamente un compromiso ontológico con las actitudes proposicionales, al saber que las redes particulares son capaces de organizar sus representaciones según categorías y que dentro de ellas hay algunas que producen una mayor activación dentro de la red, junto con el hecho de tener la capacidad de entregar más energía a una determinada red y en general por el nivel de complejidad que

18 Las licencias *creative commons*, o cc, son un tipo de licencias para contenido que pretenden facilitar la publicación y utilización de ellos, permitiendo a los usuarios compartir sus creaciones escogiendo ellos mismos las reglas que deben seguirse para el uso de ellas. Se originan como respuesta a las licencias privadas que dificultan la reutilización del contenido y las leyes de *copyright*. Para más información se puede acceder al sitio <http://www.creativecommons.cl>

19 Es otro tipo de licencia, pero se aplica al *software*, fue desarrollada por la *free software foundation* para promover el uso y creación de *software* libre. Para más información se puede ingresar a <http://www.gnu.org/licenses/gpl.html>

20 En Dennett (1981).

alcanzaría la imagen del mundo que podría llegar a interiorizar una red como la mencionada, me parece que negar que la red se comporta de manera que es posible de ser explicada según la descripción intencional sería por fuerza del chauvinismo mentalista-misteriano, que niega la capacidad de esto a la red no por sus propiedades, sino por la fuerza de la costumbre y una, casi patológica, negativa a asumir las propiedades y funciones mentales como algo aprehensible, y en última instancia imitable. Un argumento en contra podría ser afirmar que esta red podría tener una pseudo-intencionalidad y no una real como la que tenemos nosotros, frente a esto, un escepticismo de mentes artificiales, propongo como solución un “siga participando” para el problema. Esta crítica va en la misma dirección de los zombies filosóficos o el escepticismo de otras mentes, son maneras de defender un lugar privilegiado para miembros del mundo, sólo en base a alguna misteriosa certeza sobre sus capacidades intelectuales y un gran autoestima²¹.

Cabe destacar que lo dicho hasta aquí no es particularmente nuevo, de hecho es más bien un compendio de lo que han dicho y vienen diciendo hace tiempo distintos autores, pero espero haber presentado de manera clara los conceptos relevantes para la comprensión de la descripción conexionista-naturalista de las RM y lo relevante de esta para el desarrollo de una psicología o incluso de una teoría de la mente, tanto por su compromiso con las ciencias naturales como por el hecho que ha conseguido librarse de la mayor parte de los problemas que se le han presentado al camino. Esto se demostraría no sólo en el argumento usado para responder a los problemas presentados por Fodor y Pylyshyn, sino que en el hecho de lo relevantes que han mostrado ser para el actual mundo de las ciencias cognitivas, donde esta clase de modelos continúa ganando adeptos y fondos, mientras los modelos de arquitecturas clásicas se van quedando atrás y lejos del foco de atención. Si bien es verdad que la popularidad no implica veracidad, me parece que la plausibilidad biológica y el conjunto de características que diferencia este tipo de modelos los presenta como la alternativa más respetable en la actualidad.

21 En Dennett (2007) puede verse una crítica a esto mismo con un experimento mental.

Bibliografía

- Churchland (1986) Neurophilosophy, Towards a unified science of the mind brain. MIT Press.
- Churchland (1987) Epistemology in the age of neuroscience, en *Journal of philosophy*, Vol. 84.
- Churchland (1989) On the nature of theories: A neurocomputational perspective, en *Mind Desing II*.
- Churchland & Sejnowski (1992) *The computational brain*, MIT Press.
- Churchland (1996) *The engine of reason, the seat of the soul*, MIT Press.
- Churchland (2002) *Brain-wise: Studies in neurophilosophy*, MIT Press.
- Dennet (1971) Intentional Systems, en *Journal of Philosophy*, Vol. 68.
- Dennett (1981) True believers: The intentional strategy and why it works, en *Intentional Stance*.
- Dennett (1987) *Intentional Stance*, MIT Press.
- Dennett (2007) A Clever Robot, en *Time*, publicación online. URL: <http://www.time.com/time/magazine/article/0,9171,1580363,00.html>
- Fodor (1985) Fodor's guide to mental representation, en *MIND*, Vol. 94.
- Fodor (1987) *Psicosemántica: el problema del significado en la filosofía de la mente*, Tecnos.
- Fodor & Pylyshyn (1988) Connectionism and cognitive architecture: A critical analysis, en_

_____ *Cognition*, No. 28.

Frith (2007) Making up the mind, Blackwell.

Haugeland (1997) Mind Design II, MIT Press.

Ramachandran y Hubbard (2001) Synesthesia -a window into perception, thought and language, en
Journal of Consciousness Studies, 8, No. 12

Ramachandran (2003) Escuchar colores, saborear formas, en *Investigación y Ciencia*, julio.

Smolensky (1989) Connectionist modeling: Neural computation / mental connections, en
Mind Design II.

Wilson & Keil (1999) The MIT encyclopedia of cognitive science, MIT Press.