

# Arquitecturas cognitivas y cerebro: hacia una teoría unificada de la cognición

## Cognitive architectures and brain: towards an unified theory of cognition

José María Ruiz Sánchez de León  
*Universidad Complutense de Madrid*  
Miguel Ángel Fernández Blázquez  
*Universidad Complutense de Madrid*

### RESUMEN

Las arquitecturas cognitivas se definen como el conjunto de componentes esenciales de un sistema que permite el análisis de sus cogniciones y sus conductas. El objetivo de este trabajo es revisar una de las arquitecturas cognitivas que resulta más plausible desde el punto de vista neuroanatómico: ACT-R es una teoría sobre el funcionamiento de la mente humana. Tras un acercamiento inicial a sus conceptos básicos, se describen sus dos niveles de computación: un nivel simbólico que contiene toda la información declarativa del sistema, y un nivel subsimbólico que está representado como un juego de procesos en paralelo. A su vez, se relacionan los módulos que componen la arquitectura con la neuroanatomía funcional del cerebro, describiendo el funcionamiento del circuito córtico-estriado-talámico.

**Palabras clave:** Modelos computacionales; arquitecturas cognitivas; simulación; psicología cognitiva; neuropsicología cognitiva.

### ABSTRACT

Cognitive architectures are defined as the group of essential components belonging to a system which allows the analysis of its cognitions and behaviors. The aim of this study is to review one of the most plausible cognitive architectures from the neuroanatomic perspective: The Adaptive Control of Thought-Rational (ACT-R) is a theory about how human mind works. Following an initial approach to its basic concepts its two computational levels are described, these are: a symbolic level, which includes declarative information; and a sub-symbolic level which is represented as a parallel set of processes. At the same time, architecture's modules are related to brain's functional neuroanatomy describing how cortico-striatal-thalamic circuit works.

**Key words:** Computational models; cognitive architectures; simulation; cognitive psychology; cognitive neuropsychology.

---

Artículo recibido/Article received: Octubre 27 2011/October 27 2011, Artículo aceptado/Article accepted: Noviembre 12 2011/November 12 2011

Dirección correspondencia/Mail Address:

José María Ruiz Sánchez de León y Miguel Ángel Fernández Blázquez. Dpto. Psicología Básica II (Procesos Cognitivos). Facultad de Psicología. Buzón 119. Universidad Complutense de Madrid. (28223 - Madrid - España). E-mail: [jm.ruiz.sdl@gmail.com](mailto:jm.ruiz.sdl@gmail.com) / [ma.fdez.blazquez@gmail.com](mailto:ma.fdez.blazquez@gmail.com)

INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH esta incluida en PSERINFO, CENTRO DE INFORMACION PSICOLOGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET y GOOGLE SCHOLARS. Algunos de sus articulos aparecen en SOCIAL SCIENCE RESEARCH NETWORK y está en proceso de inclusion en diversas fuentes y bases de datos internacionales.

INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH is included in PSERINFO, CENTRO DE INFORMACIÓN PSICOLÓGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET and GOOGLE SCHOLARS. Some of its articles are in SOCIAL SCIENCE RESEARCH NETWORK, and it is in the process of inclusion in a variety of sources and international databases.

Hace algunos años, Newell (1990) propuso que la Neurociencia Cognitiva no podría soportar siempre la gran cantidad de modelos que, década tras década, se irían proponiendo en las diferentes áreas de conocimiento que la componen. Los resultados tienden a ser encapsulados en una u otra especialidad y, allí, se van acumulando sin que parezca existir una meta común. Así, sugirió la idea de crear una teoría unificada de la cognición que pretendiera explicar con detalle todos los mecanismos implicados en producir cualquier conducta humana. Una teoría que pudiera reproducir y explicar todos los datos que arroja la investigación. Las arquitecturas cognitivas emergieron para dar respuesta a esta necesidad de integración del conocimiento. Arquitectura (del latín, architectūra), amén de significar al arte de proyectar y construir edificios, hace referencia a la estructura lógica y física de los componentes de un sistema. Por otro lado, cognitiva, como perteneciente o relativo a las cogniciones (del latín, cognitio, -ōnis), se refiere a la acción o efecto de conocer. Así, en sentido estricto, una arquitectura cognitiva es la estructura lógica y física de los componentes de un sistema que conoce.

Sun (2002, 2004) define las arquitecturas cognitivas como el conjunto de componentes esenciales de un sistema que permite el análisis exhaustivo, en múltiples niveles y en múltiples dominios de sus cogniciones y sus conductas. Este conjunto de componentes proporciona un marco conceptual que facilita la modelización detallada de los fenómenos cognitivos a través de la división en módulos y la especificación de las relaciones entre ellos. Con ello, el análisis se realiza, principalmente a nivel computacional (Sun, 1999).

Las arquitecturas cognitivas son, en general, herederas del solucionador general de problemas que desarrollaron en los años 60 Newell y Simon (1963) y, como se expone en el presente trabajo, dan cuenta de los mecanismos que sustentan gran cantidad de conductas humanas. Se suelen basar en la idea central de regla de producción de Newell y Simon (1972). Una regla de producción es una estructura de conocimiento que consta de dos partes diferenciadas: una condición, que puede ser externa o interna al sistema, y una acción, que puede ser una respuesta motora o una operación mental. Suelen tomar la forma “SI condición ENTONCES acción”.

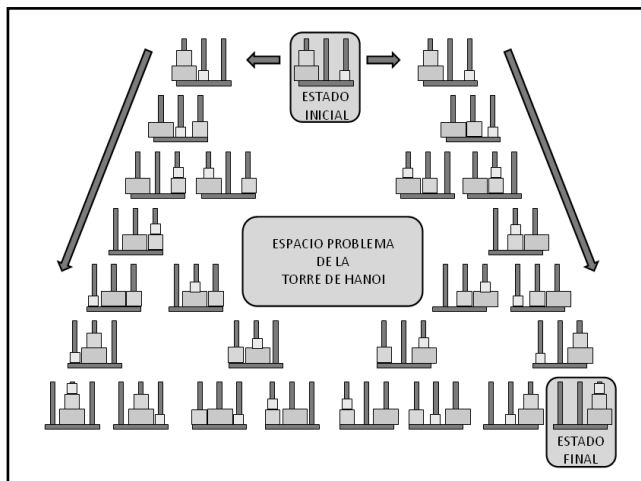
Un aspecto importante de las arquitecturas cognitivas es que consideran la conducta inteligente de los sistemas enmarcada dentro del modelo computacional de espacios problema. Así, cuando un sistema se enfrenta con una determinada tarea, ésta se convierte en un espacio problema que se identifica como una meta a resolver. Sin embargo, para solucionar un espacio problema el sistema debe, además, conocer el funcionamiento de todas sus reglas de producción, la consecuencia de sus acciones e incluso decidir qué regla utilizar en un determinado

momento entre dos posibles candidatas. Para solucionar esto se utilizan espacios multiproblema. Ciertos problemas en la resolución de una tarea se convierten en submetas de la meta original. Estas nuevas submetas no se abandonan hasta que se resuelven y entonces, se devuelve de nuevo el control a la que era meta principal.

Existen varios conceptos fundamentales que merecen la pena desatacar. Por estado, entendemos la estructura de datos que define una situación en el camino para alcanzar un objetivo. Así, para la resolución de una tarea se suele hablar de un estado inicial, estados intermedios y estado final, que es la meta. Por operador, se entiende la función que transforma un estado en otro. Un operador es la acción que resulta de la aplicación de una regla. El espacio problema es el conjunto de estados y operadores que están disponibles para alcanzar el objetivo. Las dimensiones del espacio problema pueden ser muy variables y dependen de la complejidad de la tarea. Así, se puede resumir toda la actividad cognitiva del sistema diciendo que se lleva a cabo aplicando operadores a estados dentro de un espacio problema para alcanzar una meta.

En la Figura 1 se muestra el espacio problema de la torre de Hanoi, donde pueden verse todos los movimientos posibles que dan lugar a todos los posibles estados intermedios entre el estado inicial y el estado final. Newell y Simon (1972) aseguran que el espacio problema de la torre de tres discos puede ser procesado íntegramente en la memoria operativa porque no excede su capacidad. Por el contrario, el procesamiento de los espacios problema de otras tareas más complejas, como el ajedrez, exceden dicha capacidad por lo que el sistema suele manejar subconjuntos (y submetas).

Figura 1. *Espacio problema de la Torre de Hanoi de tres discos*



Se han propuesto diferentes arquitecturas cognitivas en las últimas décadas. Algunas de ellas son la BDI (Bratman 1987), CAP2 (Schneider y Detweiler, 1987), CAPS (Just y Carpenter, 1992), 3T (Bonasso, Firby, Gat, Kortenkamp, Miller y Slack, 1997), ICARUS (Shapiro y Langley, 1999), REMI (Schooler, Shiffrin y Raaijmakers, 1999), ROUSE (Huber, Shiffrin, Lyle y Ruys, 2001) y CLARION (Sun, 2001, 2002). Sin embargo, las arquitecturas SOAR (Laird, Rosenbloom y Newell, 1987), EPIC (Meyer y Kieras, 1997) y ACT-R (Anderson y Lebiere, 1998), que se desarrolla a continuación, son las que más respaldo han obtenido.

### LA ARQUITECTURA ACT-R

Anderson y Bower (1973) desarrollaron el modelo HAM, una propuesta para considerar la memoria declarativa un conjunto de proposiciones. Este es el modelo predecesor del primer ACT (Anderson, 1976, 1982) que también implementaba un sistema de producciones en la memoria procedimental. En tercer lugar, apareció la arquitectura ACT\* (Anderson, 1983, 1989, 1992), en la que combinaban procesos simbólicos, como especificaciones formales del contenido de la memoria declarativa, con procesos subsimbólicos, que explicaban la activación en dicha memoria declarativa y su relación con la procedimental. El modelo ACT-R (de Adaptive Control of Thought - Rational) de Anderson (1993) optimizó más tarde los componentes subsimbólicos para dar cuenta de las demandas del medio, adaptando la arquitectura cognitiva a los fenómenos estocásticos del mismo. En los últimos años, las diferentes versiones de la arquitectura ACT-R (Anderson, Bothell, Byrne, Douglass, Lebiere y Qin, 2004; Anderson y Lebiere, 1998) conciben un sistema de módulos y buffers localizados en regiones cerebrales determinadas.

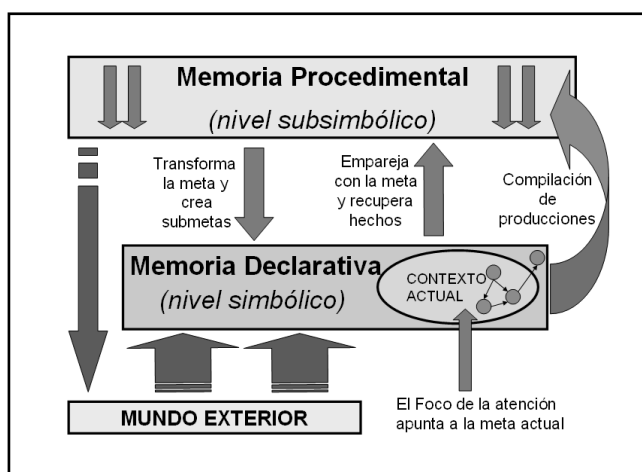
La teoría ACT, en general, viene acumulando evidencias en múltiples campos de la psicología cognitiva, desde la atención y la percepción (con modelos para dar cuenta de la búsqueda visual, los movimientos oculares o del efecto Stroop), el aprendizaje y la memoria (adquisición de destrezas, recuerdo de lista de palabras, categorización), memoria operativa, toma de decisiones y resolución de problemas (modelizando la ejecución de la torre de Hanoi, problemas matemáticos o sistemas dinámicos), lenguaje (explicando el funcionamiento de las analogías, las metáforas o el procesamiento léxico), desarrollo cognitivo, diferencias individuales, motivación y emoción.

En última instancia, ACT-R es una teoría sobre cómo trabaja la mente humana. Para los que se acercan por primera vez a sus supuestos, parece un lenguaje de programación informático. Sin embargo, es mucho más que eso. ACT-R es un marco conceptual que proporciona las herramientas necesarias para generar hipótesis y crear modelos (o programas, si se quiere) que, una vez escritos y puestos a prueba, permiten conocer mejor cómo funciona el

sistema cognitivo. Esto es así en la medida en que se comparan los resultados del programa informático con los resultados obtenidos en investigación psicológica humana, ya sean tiempos de reacción, precisión en las respuestas o registros en la resonancia magnética funcional.

ACT-R es una arquitectura cognitiva híbrida (Figura 2). Esto es, supone un nivel simbólico, que contiene toda la información declarativa del sistema, y un nivel subsimbólico, que está representado como un juego de procesos en paralelo similares a los de un modelo conexionista, que controlan muchos de los procesos simbólicos, y dan cuenta de los mecanismos de aprendizaje.

Figura 2. Estructura cognitiva de la arquitectura ACT-R



#### Nivel simbólico

En el nivel simbólico el conocimiento declarativo se sustenta por *chunks*, estructuras que representan la información a modo de proposiciones y que están compuestas por un número determinado de ranuras etiquetadas (*slots*) en las que se almacena dicha información. Ésta puede ser un valor, un hecho, antiguas metas del sistema, la meta actual e incluso información perceptiva. Cada chunk es de un tipo particular que determina su nombre y el número de sus ranuras. Toman la forma:

```
(TIPO-DE-CHUNK  
NOMBRE  
RANURA1  
RANURA2  
RANURA... n)
```

Un ejemplo de chunk meta podría ser el siguiente:

```
(META X  
ES-UNA SUMA
```

SUMANDO1 TRES  
SUMANDO2 CUATRO  
RESPUESTA VACIO)

Donde SUMANDO1, SUMANDO2 y RESPUESTA son ranuras de la META X y TRES y CUATRO son contenidos de las ranuras que hacen referencia a otros chunks de la memoria declarativa. La ranura RESPUESTA obtiene un valor de VACIO, queriendo indicar que en este caso la respuesta es desconocida por el sistema y además es una meta a resolver. Una vez resuelto el problema, el conocimiento quedaría así:

(TIPO-DE-CHUNK  
SUMAR  
SUMANDO1  
SUMANDO2  
RESPUESTA)  
(TIPO-DE-CHUNK  
NUMERO-ENTERO  
VALOR)  
(SUMAR (META X  
ES-UNA SUMA  
SUMANDO1 TRES  
SUMANDO2 CUATRO  
RESPUESTA SIETE)  
(TRES  
ES-UN NUMERO-ENTERO  
VALOR 3)  
(CUATRO  
ES-UN NUMERO-ENTERO  
VALOR 4)  
(SIETE  
ES-UN NUMERO-ENTERO  
VALOR 7))

Por otro lado, el conocimiento procedimental en el nivel simbólico consiste en reglas de producción (o producciones). El sistema de producción ACT original de Anderson (1976) suponía que un conjunto de producciones actuaba sobre una base de datos sobre hechos en la memoria declarativa.

Las producciones en ACT-R son reglas que especifican una contingencia cognitiva, es decir, cuándo una operación cognitiva debe tener lugar. Contienen una condición que especifica la circunstancia bajo la cual debe operar y una acción que se realizará en caso de que la circunstancia previa se cumpla. En concreto, las condiciones contienen patrones que deben ser comparados con la meta actual del sistema y otros chunks de la memoria declarativa. Las acciones modifican el valor de las ranuras en la meta o en otros chunks, varían la meta actual o crean submetas de la misma.

Las producciones tienen un gran nivel de abstracción, asimetría condicional y están orientadas a metas. Como ya se ha dicho, una producción se relaciona con un chunk de tipo *meta* de la memoria declarativa en el que se ha centrado el foco de la atención. Las metas se organizan en pilas de metas en las que éstas pueden ser almacenadas (*pushed*) y después ser recuperadas (*popped*).

El nivel simbólico de la arquitectura ACT-R opera en ciclos discretos deterministas. Así, al comienzo de cada ciclo, las condiciones de cada producción se comparan con las ranuras del chunk meta actual. Las producciones que se emparejan con éxito entran a formar parte del juego de producciones que entran en conflicto en ese momento. Entonces, el sistema selecciona una producción de entre éstas y consulta las condiciones que especifica recuperándolas de la memoria declarativa. Si dicha recuperación resulta total y positiva (esto es, que todas las condiciones de la producción se emparejan con las ranuras de la meta actual) la acción que contiene la producción se ejecuta, pudiendo tener como resultado una variación de la meta en vigor, el almacenamiento de ésta en la pila de metas (*push*) o la recuperación de una meta anterior ya almacenada en la pila (*pop*). Si por el contrario esa producción no se empareja con el estado y con la meta, el sistema pasa a la siguiente producción candidata del juego de producciones en conflicto.

Siguiendo con un ejemplo similar al anterior, el sistema, ante el problema de restar el número 4 del número 7, puede utilizar y seleccionar de entre sus producciones la que sigue:

SI  
LA META ES RESTAR  
*num2 A num1*  
Y LA RESPUESTA ESTÁ VACÍA  
Y EXISTE UN CHUNK DE TIPO SUMAR  
*num2 MAS num3 ES IGUAL A num1*  
ENTONCES  
PONER *num3*  
EN LA RANURA RESPUESTA DE LA META

Donde *num1* es el valor 7, *num2* es el valor 4 y *num3* el valor 3. Así, las variables que contienen las producciones son de crucial importancia para el sistema. De esta manera, esta regla es capaz de resolver cualquier resta si existe el chunk suma en la memoria declarativa.

*Nivel subsimbólico*

A diferencia del nivel simbólico, que funciona como los típicos sistemas de producción (como el primer ACT, Anderson, 1976), el nivel subsimbólico opera de manera similar a las estructuras conexionistas. Esto se debe a que el funcionamiento del nivel simbólico arroja dos

cuestiones que debían ser resueltas (Anderson, 1993; Anderson y Lebiere, 1998). ¿Cómo se ordenan las producciones que han emparejado en el juego de producciones en conflicto? Y, en el caso de que los chunks de varias producciones emparejen con la recuperación de la memoria declarativa, ¿cuál se selecciona? Además, el nivel subsimbólico resuelve y da cuenta de otros problemas como los errores del sistema y predice las latencias en las decisiones.

En primer lugar, si varios chunks de la memoria declarativa se activan ante las condiciones de varias producciones y satisfacen la comparación, el chunk más activo es el que se recupera. La activación de un chunk ( $A_i$ ) es la suma de su nivel basal de activación ( $B_i$ , entendido como la utilidad que tuvo en el pasado) y su activación asociativa (relevancia para el contexto general).

Así, ante una nueva meta de tipo restar el valor 2 al valor 8, los chunk que contienen los valores 2 y 8 incrementan su actividad por ser relevantes en el contexto. A su vez, si existe el chunk de tipo suma con la información  $2+6=8$ , éste también recibirá una atención extra ya que, utilizando la regla de producción del ejemplo anterior, el problema se resolverá rápidamente.

El nivel basal de activación  $B_i$  del chunk  $C_i$  es una estimación independiente del contexto de la probabilidad de que  $C_i$  empareje con una producción. El parámetro  $B_i$  está determinado por la frecuencia con la que se ha usado  $C_i$  y con la recencia de  $C_i$  en el momento dado. La activación asociativa, por otro lado, es dependiente del contexto y se refiere al sumatorio del producto de las fuentes de activación ( $W_j$ , que son los componentes de la meta actual que se relacionan con el contexto) y la fuerza asociativa de todos los chunks relativos activados ( $S_{ji}$ ). En ACT-R las fuentes de activación hacen referencia a la atención que se presta a los elementos, por lo que es de capacidad limitada (e introduce también una variable de diferencias individuales). Se añade una variable de ruido ( $N(0,s)$ ) a cada activación imprimiendo un carácter estocástico, con dos ruidos compuestos: uno permanente que da cuenta de la variabilidad en la codificación y otro transitorio que depende del momento. Así, la activación de un chunk ( $i$ ) resulta:

$$A_i = B_i + \sum_j W_j \cdot S_{ji} + N(0, s)$$

No obstante, puede darse el caso de que los chunks se emparejen solo parcialmente con el patrón de recuperación. En ese caso pueden ser elegidos, pero su nivel de activación es penalizado en una cantidad proporcional al grado de desajuste entre el patrón y los

valores actuales del chunk. Por ello, la función final que se refiere al nivel de activación del chunk queda como sigue:

$$A_i = B_i + \sum_j W_j \cdot S_{ji} + \sum_k MP_k \cdot Sim_{kl} + N(0, s)$$

Donde  $MP_k$  es la penalización por no emparejar y  $Sim_{kl}$  el valor de similaridad actual. En el caso de que la activación de ningún chunk alcance el umbral de recuperación, dicha recuperación fracasa (obligando al sistema a crear submetas de la original).

Así, el tiempo que invierte el sistema en recuperar un chunk de la memoria ( $Time_i$ ) es una función ( $F$ ) con una caída exponencial de su nivel de activación previo ( $A_i$ ). A menos activación del chunk, más tiempo se invierte en recuperarlo.

$$Time_i = F \cdot e^{-f(A_i + S_p)}$$

Como un chunk siempre es recuperado por una regla de producción, esta ecuación representa el tiempo en recuperar el chunk ( $i$ ) por la producción ( $p$ ). La fuerza de la regla de producción ( $S_p$ ) juega también un papel importante en el tiempo que tarda en recuperar un determinado chunk. Los parámetros  $F$  y  $f$  son fijos en ACT-R con un valor por defecto igual a 1. Para evitar tiempos de recuperación superiores a un segundo se establece un umbral constante que, en caso de sobrepasarse, hace que el chunk ( $i$ ) no sea recuperado.

Un segundo hecho relevante en el nivel subsimbólico es que las producciones se seleccionan en orden descendente de utilidad o ganancia esperada. Por un lado, se asigna una utilidad o ganancia ( $G$ ) al chunk que es la meta actual del sistema que se corresponde con lo que merece la pena satisfacerla. Por otro, a cada producción ( $p$ ) se le asocia una probabilidad de éxito ( $P_p$ ) y un coste ( $C_p$ ) por conseguir resolver la meta a la que se dedica. La unidad de coste en ACT-R es el tiempo. La utilidad o ganancia esperada ( $EG_p$ ) de una producción es el resultado de la probabilidad de éxito por ganancia del chunk meta menos los costes de la producción. Así resulta:

$$EG_p = P_p \cdot G - C_p$$

Donde la probabilidad de éxito de la producción es igual a:

$$P_p = \frac{\text{Éxitos}}{\text{Éxitos} + \text{Fracasos}}$$

En una versión más avanzada, el sistema introduce además una variable de ruido a la utilidad esperada de la producción (haciendo que la selección de las producciones sea estocástica). Así, la probabilidad de que una producción ( $P_p$ ) sea elegida es:

$$P_p = \frac{e^{\frac{\text{Éxitos } p}{t}}}{e^{\frac{\text{Éxitos } p}{t}} + e^{\frac{\text{Fracasos } p}{t}}}$$

Donde el parámetro ( $t$ ) refleja el ruido del medio en la evaluación de las producciones. Así, mientras que el modelo ACT-R tiene un algoritmo serial preciso en su nivel simbólico, en su nivel subsimbólico ha evolucionado hacia una predominancia de la estocasticidad y la continuidad de la ejecución humana (Anderson *et al.*, 2004).

#### Aprendizaje en ACT-R

Así como la arquitectura ACT-R propone dos sistemas de memoria diferentes, cada uno de ellos con dos niveles de descripción, también propone diferentes mecanismos de aprendizaje. Dichos mecanismos de aprendizaje dentro del nivel simbólico especifican cómo se añaden los chunks nuevos a la memoria declarativa y cómo las reglas lo hacen a la memoria procedimental. Por otro lado, en el nivel subsimbólico da cuenta de cómo cambian los valores de los parámetros. Como en la arquitectura SOAR (Laird *et al.*, 1987), nunca se borra conocimiento de la memoria aunque éste pueda llegar a ser prácticamente irrecuperable de la misma.

Un chunk nuevo tiene la posibilidad de acceder a la memoria declarativa por dos medios, bien porque haya sido un objeto perceptivo externo, o bien porque sea un chunk creado por el sistema a partir del procesamiento y el logro de metas por parte de la arquitectura. En este segundo caso, los chunks creados dentro del sistema son siempre antiguas metas, como en el ejemplo del resultado de la suma anterior. Por ello, Anderson (1993) afirma que cualquier chunk de la memoria declarativa que no ha accedido al sistema por vías perceptivas, en algún momento fue una meta actual. Así, el aprendizaje se produce en la medida en que el sistema no necesita reprocesar la suma 3+4 desde el punto en que la ranura RESPUESTA tenía un valor VACÍO, sino que recuperará directamente el chunk en el que dicha ranura ya tiene el valor SIETE.

El aprendizaje de nuevas reglas de producción es un proceso mucho más complicado. Las producciones se aprenden a partir de ejemplos que se almacenan en un tipo especial de chunk llamado *dependencia* (*dependency chunk*). Cuando se resuelve una meta por medio de un proceso que implica varias reglas se crea un chunk de dependencia que contiene información acerca de cómo se

resolvió, qué chunks fueron recuperados y cuáles fueron las submetas creadas para tal efecto.

Entonces, cuando el sistema ACT-R se enfrenta a una situación similar (con una meta similar), el chunk de dependencia con ese conocimiento es recuperado (“*popped*”) y automáticamente se compilada una nueva producción para solucionar el problema mediante un mecanismo de analogía. Así, el sistema resuelve mediante esa generalización la meta actual en un solo paso en lugar de reprocesar de nuevo todas las submetas que en su momento utilizó. Es importante destacar que, como el chunk de dependencia no es un chunk perceptivo, en algún momento fue una meta del sistema y, por ello, se dice que el aprendizaje de reglas de producción es dependiente de la memoria declarativa y, además, un proceso orientado a metas.

Por otro lado, en el nivel subsimbólico, el aprendizaje se basa en reglas bayesianas para ajustar las estimaciones de los parámetros en función de la experiencia (Berger, 1985). De acuerdo con estos principios, una nueva estimación de un determinado parámetro se basa en un valor previo y en la experiencia actual. Así, el nivel basal de activación de un chunk ( $B_i$ ) como ya se ha comentado anteriormente es una estimación de la necesidad de usar dicho chunk independientemente del contexto. Si el chunk ha sido recuperado muchas veces en el pasado, la probabilidad de volver a recuperarlo aumenta. Por el contrario, si ha sido recuperado pocas veces, su probabilidad de ser usado es pequeña.

Por ello, cada vez que un chunk es recuperado, su nivel basal de activación debería subir (y cada vez que no se usa, bajar). Eso es lo que hace exactamente el mecanismo de aprendizaje del nivel subsimbólico, aumentar dicho valor si se usa el chunk habitualmente y disminuirlo cuando pasa tiempo sin usarse. La fórmula que calcula el nivel basal de activación en un momento ( $t$ ) es:

$$B_i(t) = \log \sum_{j=1}^n (t - t_j)^{-d}$$

Donde ( $n$ ) es el número de veces que se ha recuperado de la memoria el chunk ( $i$ ), y ( $t$ ) representa el momento en el que tiene lugar cada recuperación. Por ello, cuanto más tiempo haga de su última recuperación, menor nivel basal de activación. El parámetro ( $d$ ) es fijo en ACT-R y representa la caída del nivel basal en la memoria declarativa (su valor por defecto es 0,5).

Neves y Anderson (1981) aportaron los datos experimentales que demuestran cómo el aprendizaje de la arquitectura ACT-R se ajusta a la clásica curva de aprendizaje asintótica, así como Bahrick (1984) lo hiciera

con la curva del olvido. Sin embargo, la aportación más relevante de la ecuación del nivel basal de activación es la relación que propone entre la variable *cantidad de práctica* y la variable *tiempo que se retiene la práctica*. Formalmente, la ecuación que engloba estos tres aspectos (curva de aprendizaje, curva de olvido y la relación cantidad-duración de la práctica) en la teoría ACT-R es:

$$\text{Ejecución} = N \cdot T^{-d}$$

Esto implica que la ejecución mejora continuamente con la práctica (N es la cantidad de práctica) y empeora, también continuamente, con el intervalo de tiempo entre prácticas (T es el tiempo sin practicar). El hecho de que estos dos factores se multipliquen quiere decir que aumentar la práctica es la manera de conservar el conocimiento frente al paso del tiempo (Anderson, 1995).

#### Neuroanatomía funcional en ACT-R.

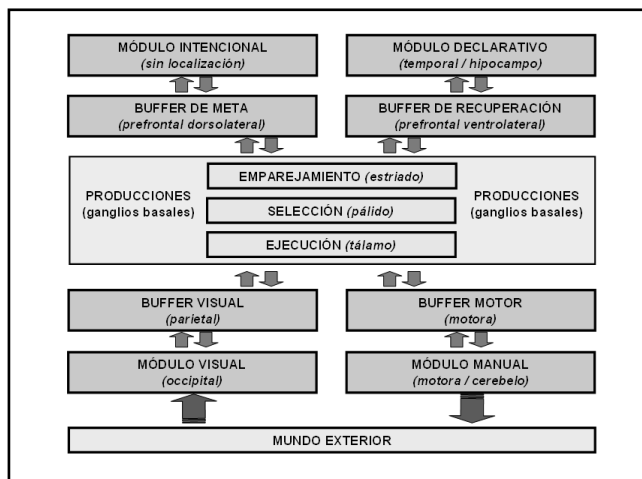
La arquitectura ACT-R, en la actualidad ACT-R 6 versión 1.4, presupone la existencia de un conjunto de módulos encargados del procesamiento de diferentes tipos de información. Desde su versión 5.0, ACT-R trabaja con varios módulos de procesamiento: un módulo visual, un módulo manual, un módulo declarativo que recupera la información de la memoria y un módulo de metas encargado de mantener las intenciones del sistema actualizadas. El sistema central de producciones se encarga de la coordinación de estos módulos durante la conducta (Anderson *et al.*, 2004).

Este sistema central no es sensible al procesamiento que se lleva a cabo en el resto del sistema, sino que responde a la información limitada que es depositada en sus *buffers*. En términos de Fodor (1983), la información está encapsulada y sólo se intercomunican mediante el tipo de información que comparten y es accesible. Así, no toda la información del campo visual es atendida en un momento dado o no todo el contenido de la memoria semántica es accesible a la conciencia a la vez, sino cuando el sistema la necesita. La arquitectura EPIC (del inglés *executive-process/interactive control*, proceso-ejecutivo/control-interactivo) de Kieras, Meyer, Mueller y Seymour (1999) adoptó un sistema ejecutivo similar. Se propone que esos procesos tienen un grado de similitud con los sistemas *esclavos* de la memoria operativa de Baddeley (1986).

El *buffer* de metas, relacionado entre otras estructuras con el cortex prefrontal dorsolateral (CPFDL), conoce el estado interno del sistema, mientras que el *buffer* de recuperación, relacionado con la corteza prefrontal ventrolateral (CPFVL), mantiene la información recuperada de la memoria declarativa (Nyberg, Cabeza y Tulving,

1996; Buckner, Kelley y Petersen, 1999; Wagner, Pare´-Blagoev, Clark y Poldrack, 2001). Esa distinción entre CPFDL y CPFVL es consistente con muchos hallazgos anteriores (Cabeza, Dolcos, Graham y Nyberg, 2002; Fletcher y Henson, 2001; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997).

Figura 3. Organización de la información en el cerebro según la arquitectura ACT-R



El buffer manual es responsable del control motor de la mano y está relacionado con las áreas corticales motoras y somatosensoriales responsables de controlar y monitorizar el movimiento de las mismas. Por otro lado, uno de los buffers visuales se encarga del procesamiento del dónde y analiza localizaciones, relacionado con la vía dorsal del procesamiento visual, y el otro con el procesamiento del qué, analiza los estímulos, relacionado con la vía ventral. Los buffers de los módulos perceptivos y motores están basados en trabajos anteriores de la arquitectura EPIC (Meyer y Kieras, 1997).

Además de todo lo anterior, la arquitectura ACT-R supone que los ganglios basales y sus proyecciones implementan en el sistema biológico las reglas de producción, de manera que las áreas corticales que conectan con el cuerpo estriado parecen encargarse del reconocimiento de patrones (Amos, 2000; Frank, Loughry y O'Reilly 2000; Houk y Wise, 1995; Wise, Murria y Gerfen, 1996; Stocco, Lebiere y Anderson, 2010). Éste proyecta sus fibras hacia el globo pálido y éste otro, a su vez, hacia el tálamo, que en última instancia selecciona las acciones de la corteza. Así, tras una determinada respuesta correcta, la activación estriatal inhibe el funcionamiento del globo pálido, que no vuelve a inhibir al tálamo, con lo que la regla de producción que desencadena dicha respuesta correcta se selecciona como respuesta dominante (Graybiel y Kimura, 1995). Se estima que al menos cinco regiones frontales reciben sus proyecciones del tálamo y están controladas por

este bucle basal. Regiones todas, que se relacionan neuropsicológicamente con el control de la conducta (Middleton y Strick, 2000). En resumen, el cuerpo estriado reconoce patrones, el globo pálido resuelve los conflictos y el tálamo controla la ejecución de las respuestas. Así, una regla de producción en ACT-R supone un ciclo de procesamiento desde la corteza hasta los ganglios basales y de nuevo a la corteza.

Por último, cabe destacar que la arquitectura asume la coexistencia de procesos seriales y paralelos. Por ejemplo, el sistema visual procesa simultáneamente todo el campo visual (procesamiento en paralelo). Sin embargo, existen dos niveles en los que se hace patente el procesamiento serial, al aparecer efectos de cuello de botella. Primero, cada buffer solo puede contener una unidad declarativa de conocimiento (chunk), con lo que sólo se puede recuperar una huella mnésica cada vez o sólo se puede codificar un objeto del campo visual. Segundo, sólo una regla de producción es seleccionada para entrar en el ciclo de procesamiento córtico-estriado-talámico.

En los últimos años, ACT-R está proporcionando resultados excelentes comparando sus predicciones con los resultados hallados mediante resonancia magnética funcional (Anderson, Fincham, Qin, y Stocco, 2008; Anderson, Qin, Stenger y Carter, 2004; Sohn, Albert, Stenger, Jung, Carter y Anderson, 2007; Sohn, Goode, Stenger, Carter y Anderson, 2003; Sohn, Goode, Stenger, Jung, Carter y Anderson, 2005; Anderson, Carter, Fincham, Ravizza, y Rosenberg-Lee, 2008). Entre otros procesos cognitivos, se han puesto a prueba la torre de Hanoi (Anderson, Albert y Fincham, 2005), diferentes operaciones matemáticas (Danker y Anderson, 2007; Qin, Carter, Silk, Stenger, Fissell, Goode y Anderson, 2004), tareas de planificación (Fincham, Carter, van Veen, Stenger y Anderson, 2002) y de cambio atencional (Sohn, Ursu, Anderson, Stenger y Carter, 2000; Altmann y Gray, 2008).

## CONCLUSIONES

Las arquitecturas cognitivas proporcionan el soporte físico (hardware) y las herramientas simbólicas (software) necesarios para describir y predecir la conducta de un sistema neurocognitivo en una situación determinada. Aplicadas a la cognición humana, las arquitecturas cognitivas ofrecen un excelente marco experimental que permite modelar y poner a prueba una determinada hipótesis sobre su funcionamiento. Durante las últimas décadas han proliferado distintos tipos de arquitecturas cognitivas. La principal diferencia entre ellas radica en los supuestos teóricos a partir de los cuales han sido desarrolladas. La mayor parte de las arquitecturas han sido diseñadas con la intención de dar respuesta a un aspecto específico del funcionamiento cognitivo. Por este motivo, es frecuente encontrar modelos dependientes de la tarea

que, lejos de explicar el funcionamiento de un determinado proceso cognitivo, tan solo son capaces de simular cómo se resuelve eficazmente dicha tarea concreta (como por ejemplo, la simulación del problema de la torre de Hanoi).

Asimismo, si bien es mucho lo que se sabe actualmente acerca de la estructura y funcionamiento de los procesos cognitivos de forma individual, se desconoce aún qué mecanismo produce la integración de todos ellos dentro del cerebro (puesto que los procesos cognitivos están íntimamente interrelacionados, parece razonable pensar que comparten determinadas redes neuronales, pero hasta el momento no se ha prestado excesiva atención a explicar cómo se produce esa interacción). Las arquitecturas cognitivas ofrecen la posibilidad de modelar cómo se relacionan los diferentes procesos cognitivos. De este modo, se convierten en herramientas poderosas para el investigador por dos razones. Por un lado, permiten pasar de un planteamiento meramente teórico a una implementación virtual de un determinado modelo cognitivo. Por otro, las arquitecturas cognitivas pueden desempeñar un importante papel como predictores de la conducta humana, tanto normal como patológica.

Con el fin de seleccionar qué tipo de arquitecturas pueden resultar de mayor utilidad para explicar la cognición humana, recientemente se han enumerado una serie de requisitos fundamentales que toda arquitectura debe cumplir. Entre tales requisitos se pueden destacar los siguientes: validez ecológica, plausibilidad biológica, ajuste del modelo al conocimiento actual de la cognición, funcionamiento modular del sistema, existencia de conexiones de reentrada masiva entre los módulos superiores e inferiores del modelo, y capacidad para aprender tanto implícita como explícitamente (Sun, 2004). ACT-R, al contrario que la gran mayoría de arquitecturas, ha demostrado reunir muchas de estas cualidades, lo que la convierte en un potente instrumento con el que modelar el comportamiento humano.

No obstante, la principal diferencia entre ACT-R y otras arquitecturas cognitivas es el enfoque integrador. ACT-R no sólo se conforma con describir el comportamiento de un sistema en una tarea determinada, sino que va más allá. En última instancia, su intención es explicar el funcionamiento de los diferentes procesos cognitivos de forma individual y, lo que es más novedoso, de forma integrada (el fenómeno de la integración de la cognición es obviado con frecuencia por el resto de arquitecturas dada su complejidad). De esta forma, ACT-R constituye un modelo más ajustado a la concepción que se tiene actualmente sobre el funcionamiento cognitivo. Por otro lado, cualquier modelo que pretenda dar cuenta de cómo se produce la integración de la cognición no puede ignorar la neuroanatomía. Una vez más, el enfoque integrador de ACT-R permite ajustar el sistema cognitivo a



los conocimientos neuroanatómicos que hoy en día se disponen, lo que dota al modelo de una mayor validez y plausibilidad biológica.

Finalmente, a pesar de que la explicación de los procesos cognitivos en el marco del conocimiento neuroanatómico actual no está aún completa ni es totalmente satisfactoria, ACT-R constituye una aproximación válida para explicar cómo se integran los procesos cognitivos en el cerebro humano. Así, la Neurociencia Cognitiva se acercará un poco más a esa teoría unificada de la cognición propuesta por Newell (1990) que integrará el conocimiento aportado por todas las disciplinas que la componen.

### REFERENCIAS

- Altmann, E.M. y Gray, W.D. (2008). An integrated model of cognitive control in task switching. *Psychological Review*, 115, 602-639.
- Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience*, 12, 505-519.
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Anderson, J. R. (1989). Practice, working memory, and the ACT\* theory of skill acquisition: A Comment on Carlson, Sullivan, and Schneider. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 527-530.
- Anderson, J. R. (1992). Automaticity and the ACT\* theory. *American Journal of Psychology*, 105, 165-180.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1995). *Learning and Memory*. New York: Wiley.
- Anderson, J.R., Albert, M.V. y Fincham, J.M. (2005) Tracing Problem Solving in Real Time: fMRI Analysis of the Subject-Paced Tower of Hanoi. *Journal of Cognitive Neuroscience*, 17, 1261-1274.
- Anderson J.R., Bothell D., Byrne M.D., Douglass S., Lebiere C. y Qin Y. (2004). An integrated theory of the mind. *Psychological Review*, 111, 1036-60.
- Anderson, J.R. y Bower, G.H. (1973). *Human associative memory*. Washington: Winston and Sons.
- Anderson, J.R., Carter, C.S., Fincham, J.M., Ravizza, S.M., y Rosenberg-Lee, M. (2008). Using fMRI to test models of complex cognition. *Cognitive Science*, 32, 1323-1348.
- Anderson, J.R., Fincham, J.M., Qin, Y., y Stocco, A. (2008). A Central circuit of the mind. *Trends in Cognitive Science*. 12, 136-143.
- Ruiz Sánchez de León, J.M., Fernandez Blázquez, M.A., (2011). Arquitecturas cognitivas y cerebro: hacia una teoría unificada de la cognición. *International Journal of Psychological Research*, 4(2), 38-47.
- Anderson, J.R. y Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.
- Anderson, J.R., Qin, Y., Stenger, V.A. y Carter, C.S. (2004). The relationship of three cortical regions to an information-processing model. *Journal of Cognitive Neuroscience*, 16, 637-653.
- Baddeley, A.D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Bahrack, H.P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 113, 1-24.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.
- Bonasso, R. P., Firby, J. R., Gat, E., Kortenkamp, D., Miller, D. P. y Slack, M. G. (1997). Experiences with an Architecture for Intelligent, Reactive Agents. *Journal of Experimental and Theoretical Artificial Intelligence*, 9, 237-256.
- Bratman, M. E. (1987). *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA.
- Buckner, R. L., Kelley, W. M. y Petersen, S. E. (1999). Frontal cortex contributes to human memory formation. *Nature Neuroscience*, 2, 311-314.
- Cabeza, R., Dolcos, F., Graham, R. y Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *NeuroImage*, 16, 317-330.
- Danker, J.F. y Anderson, J.R. (2007). The roles of prefrontal and posterior parietal cortex in algebra problem-solving: A case of using cognitive modeling to inform neuroimaging data. *Neuroimage*, 35, 1365-1377.
- Fincham, J.M., Carter, C.S., van Veen, V., Stenger, V.A. y Anderson, J.R. (2002). Neural mechanisms of planning: A computational analysis using event-related fMRI. *Proceedings of the National Academy of Sciences*, 99, 3346-3351.
- Fletcher, P.C. y Henson, R.N.A. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain*, 124, 849-881.
- Fodor, J.A. (1983). *The modularity of the mind*. Cambridge, MA: MIT Press/Bradford Books.
- Frank, M.J., Loughry, B. y O'Reilly, R.C. (2000). *Interactions between frontal cortex and basal ganglia in working memory: A computational model* (Tech. Rep. No. 00-01). Boulder: University of Colorado, Institute of Cognitive Science.
- Graybiel, A.M. y Kimura, M. (1995). Adaptive neural networks in the basal ganglia. En J.C. Houk, J.L. Davis y D.G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 103-116). Cambridge, MA: MIT Press.
- Houk, J.C. y Wise, S.P. (1995). Distributed modular architectures linking basal ganglia, cerebellum, and

- cerebral cortex: Their role in planning and controlling action. *Cerebral Cortex*, 2, 95-110.
- Huber, D. E., Shiffrin, R. M., Lyle, K. B. y Ruys, K. I. (2001). Perception and preference in short-term word priming. *Psychological Review*, 108, 149-182.
- Just, M.A. y Carpenter, P.A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99, 122-149.
- Kieras, D.E., Meyer, D. E., Mueller, S. y Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor performance. En P. Shah y A. Miyake (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 183-223). Cambridge, England: Cambridge University Press.
- Laird, J.E., Newell, A. y Rosenbloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, 33, 1-64.
- Meyer, D.E. y Kieras, D.E. (1997). A computational theory of executive cognitive processes and multiple-task performance. Part 1. Basic mechanisms *Psychological Review*, 104, 2-65.
- Middleton, F.A. y Strick, P.L. (2000). Basal ganglia and cerebellar loops: Motor and cognitive circuits. *Brain Research Reviews*, 31, 236-250.
- Neves, D.M. y Anderson, J.R. (1981). Knowledge compilation: mechanisms for the automatization of cognitive skills. En J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57-84). Hillsdale: Erlbaum.
- Newell, A. (1990). *Unified theories of cognition*. Harvard University Press, Cambridge, MA.
- Newell, A. y Simon, H. (1963). GPS, a program that simulates human thought. En E.A. Feigenbaum y J. Feldman (Eds.), *Computers and Thought*. New York: McGraw-Hill.
- Newell, A. y Simon, H. A. (1972). *Human problem solving*. Prentice Hall, Englewood Cliffs, NJ.
- Nyberg, L., Cabeza, R. y Tulving, E. (1996). PET studies of encoding and retrieval: The HERA model. *Psychonomic Bulletin & Review*, 3, 135-148.
- Qin, Y., Carter, C. S., Silk, E., Stenger, V. A., Fissell, K., Goode, A. y Anderson, J.R. (2004). The change of the brain activation patterns as children learn algebra equation solving. *Proceedings of National Academy of Sciences*, 101, 5686-5691.
- Schneider, W. y Detweiler, M. (1987). A connectionist/control architecture for working memory. En G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 53-119). San Diego, CA: Academic Press.
- Schooler, L., Shiffrin, R.M. y Raaijmakers, J.G.W. (2001) A model for implicit effects in perceptual identification. *Psychological Review*, 108, 257-272.
- Shapiro, D. y Langley, P. (1999). Controlling physical agents through reactive logic programming. *Proceedings of the Third International Conference on Autonomous Agents* (pp. 386-387). Seattle: ACM Press.
- Sohn, M.H., Albert, M.V., Stenger, V.A., Jung, K.J., Carter, C.S. y Anderson, J.R. (2007). Anticipation of conflict monitoring in the anterior cingulate cortex and the prefrontal cortex. *Proceedings of National Academy of Science*, 104, 10330-10334.
- Sohn, M.H., Goode, A., Stenger, V. A, Carter, C. S. y Anderson, J.R. (2003). Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex. *Proceedings of National Academy of Sciences*, 100, 7412-7417.
- Sohn, M.H., Goode, A., Stenger, V.A, Jung, K.J., Carter, C.S. y Anderson, J.R. (2005). An information-processing model of three cortical regions: Evidence in episodic memory retrieval. *NeuroImage*, 25, 21-33.
- Sohn, M.H., Ursu, S., Anderson, J.R., Stenger, V.A. y Carter, C.S. (2000). The role of prefrontal cortex and posterior parietal cortex in task-switching. *Proceedings of National Academy of Science*, 97, 13448-13453.
- Stocco, A., Lebiere, C., y Anderson, J.R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological Review*, 117, 540-574.
- Sun, R. (1999). Accounting for the computational basis of consciousness: a connectionist approach. *Consciousness and Cognition*, 8, 529-565.
- Sun, R. (2002). *Duality of the mind*. Mahwah, NJ: Erlbaum.
- Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology*, 17, 341-373
- Thompson-Schill, S.L., D'Esposito, M., Aguirre, G.K. y Farah, M.J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A re-evaluation. *Proceedings of the National Academy of Science*, 94, 14792-14797.
- Wagner, A.D., Pare´-Blagoev, E.J., Clark, J. y Poldrack, R.A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, 31, 329-338.
- Wise, S.P., Murray, E.A. y Gerfen, C. R. (1996). The frontal cortexbasal ganglia system in primates. *Critical Reviews in Neurobiology*, 10, 317-356.