

Memoria de trabajo, atención y composicionalidad

Pedro J. Ramos
Josep M. Sopena
Elizabeth Gilboy
Universitat de Barcelona

Es una práctica habitual en Psicología desarrollar microteorías para explicar datos muy concretos obtenidos en un conjunto limitado de tareas. Cuando todo el peso de la elaboración teórica recae en unos datos empíricos específicos sin tener en cuenta otras restricciones teóricas más generales, los grados de libertad para elaborar un modelo son en general muchos y esto hace que la proliferación de modelos sea considerable sin que se disponga de las suficientes restricciones para descartar cuáles son inadecuados. Las discusiones sobre qué modelos son los adecuados se enquistan en polémicas interminables. Teorías más generales permitirían integrar un mayor número de datos (empíricos y computacionales) en un todo con sentido y así se podrían descartar, de forma más simple y efectiva, los modelos que no se ajustan a estas restricciones. En este trabajo hemos intentado integrar en una todo con sentido algunas de las teorías existentes sobre un conjunto amplio de procesos (memoria de trabajo, razonamiento, procesamiento visual, procesamiento de lenguaje) y una amplia variedad de datos psicológicos, neurofisiológicos y computacionales. Como resultado se obtiene una teoría más general que, de forma razonable, permite imponer un amplio conjunto de restricciones a la hora de elaborar modelos sobre ciertos fenómenos y permite entre otras cosas descartar como inadecuados modelos actuales que gozan de cierto prestigio.

Palabras clave: memoria de trabajo, atención, binding, composicionalidad.

Working memory, attention and compositionality

It is standard practice in Psychology to develop microtheories to explain very specific data obtained from a reduced set of tasks. If more general theoretical constraints are disregarded, the results may be an uncontrolled proliferation of models. Discussions of which models are appropriate and which are not cause

endless controversy. With a more general theory, one could put together a larger body of data (both empirical and computational) into a meaningful whole and could reject, simply and affectively, the models that do not meet the theory's constraints. In this paper we bring together some of the existing theories on a set of processes (working memory, reasoning, visual and language processing) and a large variety of psychological, neurophysiological and computational data. We thus formulate a more general theory which, in a reasonable way, imposes a larger set of constraints on the preparation of models and challenges the validity of some inappropriate models that currently enjoy a certain prestige.

Keywords: working memory, attention, binding, compositionality.

A diferencia de lo que ocurre en otras disciplinas científicas más avanzadas en las que se han desarrollado algunas explicaciones lo suficientemente generales o integradoras como para poder sintetizar una enorme y diversa cantidad de datos empíricos en un todo con sentido, una práctica habitual en Psicología consiste en desarrollar microteorías para explicar sólo una tarea o un fenómeno específico. Las microteorías se desarrollan por un proceso ad-hoc de tipo *bottom-up*: los datos empíricos de un tipo de tareas experimentales concretas determinan por completo la teoría. Se les denomina teorías, pero de hecho hay muy poca elaboración teórica ya que la 'teoría' se limita a describir los resultados. Mientras que si se dispone de una teoría integradora más general, la aplicación "local" de la teoría a un fenómeno concreto no requiere poner todo el peso de la elaboración teórica en la interpretación de unos datos experimentales concretos, ya que las restricciones que impone la teoría y su caracterización (aunque sea parcial) son *previas* a la disponibilidad de esos datos. Lo que permiten los datos es confirmar las predicciones de la teoría o completarla y si las predicciones no se cumplen modificarla parcialmente o cuestionarla más ampliamente. Lo que no deberían permitir unos datos obtenidos en una sola tarea específica o en un solo tipo de tareas es "generar" por completo la "teoría".

Las restricciones computacionales se habrían de tener especialmente en cuenta. Serían las que, por ejemplo, nos permitirán entender la razón por la cual las tareas de razonamiento o de procesamiento del lenguaje requerirían que los sentidos o los conceptos estuvieran organizados de una determinada forma. Parece que existe un cierto acuerdo en aceptar que la mejor forma de entender el cerebro es concebirlo como un sistema de cómputo. Para muchos es la única forma.

Sin embargo, los modelos computacionales gozan de cierto desprestigio por razones similares a las expuestas anteriormente. En general se trata de micromodelos que han sido elaborados con el único objetivo de ajustarse a unos resultados específicos de una tarea específica por un procedimiento *bottom-up* ignorando otro tipo de restricciones. En la inmensa mayoría de los casos estos micromodelos se han probado con lo que se denomina en Inteligencia Artificial (IA) "problemas de juguete" (consistentes en un número reducido de ejemplos muy fáciles de resolver) o no se han llegado a implementar nunca; no tratan problemas reales y la mayoría no podría hacerlo. Se da la circunstancia de que siendo las restricciones tan pocas, existen varios modelos computacionales distintos que partiendo de presupuestos muy diferentes (por ejemplo,

modularidad o interactividad) explicarían por igual los datos. Como señala Newell (1990), los grados de libertad son demasiados y manipulando adecuadamente los numerosos parámetros libres que poseen los modelos es muy fácil ajustarse a los datos. Por otra parte, la noción de computación, tal como se ha utilizado por la ortodoxia chomskiana, ha contribuido también notablemente a este desprestigio. En este caso se ignora si la propuesta, considerada computacional, es tratable o no computacionalmente. Es conocido que las teorías chomskianas son extremadamente complejas y que hasta ahora no se han podido implementar en un modelo tratable computacionalmente (Berwick y Weinberg, 1982; Sopena y Lloberas, en preparación). Este hecho origina que la palabra “computacional” quede vacía de contenido.

La falta de las restricciones que impondría una teoría más general que fuera computable origina que el número de micromodelos posibles que se pueden elaborar para un determinado fenómeno sea considerable y que las discusiones sobre qué micromodelo es el adecuado se enquisten con el paso de los años sin que se pueda llegar a una solución razonable y sin que, por más datos experimentales nuevos que se obtengan o por más simulaciones que se llevan a cabo, se pueda resolver la situación. Claros ejemplos de lo que acabamos de decir serían las polémicas originadas en el estudio de los fenómenos del *priming* semántico, el acceso al léxico, o la ambigüedad sintáctica.

En este artículo seguimos la sugerencia de Newell para establecer el tipo de restricciones que una teoría del *priming*, del acceso léxico o de la desambiguación sintáctica debería plantearse e intentamos definir (aunque por problemas de espacio no podremos detallar) una teoría más general unificada desde la que observar estos fenómenos. Hemos tenido en cuenta el mayor número de restricciones que nos ha sido posible considerar. Estas restricciones incluyen las características de la memoria de trabajo y los procesos atencionales y su realización neurofisiológica. También la consideración de cuestiones computacionales como la composicionalidad. El hecho de tener en cuenta todas estas restricciones para elaborar una teoría más general no garantiza que la teoría sea la correcta pero sí que, a nuestro parecer, permite acercarnos de forma más adecuada al fenómeno en estudio y permite establecer nuevas estrategias de investigación más prometedoras que la estrategia que se sigue actualmente basada en una confianza excesiva en los datos empíricos obtenidos (como si solamente con ellos pudiéramos dar respuesta a todos los problemas) y en un grado de elaboración teórica que apenas se “despega” de los datos. Empezaremos analizando las restricciones que provendrían de las características de la memoria de trabajo.

La memoria de trabajo

La memoria de trabajo (MT) se suele caracterizar como un espacio de trabajo mental, una especie de pizarra que permite almacenar de forma temporal una reducida cantidad de información para manipularla mientras se lleva a cabo una tarea cognitiva (Baddeley, 1986; Cowan, 2005; Miyake y Shah,

1999). Se considera que la MT juega un papel clave en procesos fundamentales como el aprendizaje, la comprensión del lenguaje, el razonamiento, la resolución de problemas, la planificación o la categorización. De hecho, es difícil pensar en alguna tarea cognitiva compleja que no requiera el uso de la MT. Parece existir un acuerdo generalizado en que sus características principales serían la de tener una capacidad limitada y la de que la información contenida en ella se halla en un estado de rápida accesibilidad. No obstante prácticamente todos los otros aspectos de su caracterización han sido motivo de una fuerte controversia (Miyake y Shah, 1999; Cowan, 2005). En la siguiente sección describiremos algunos de los modelos sobre la arquitectura de la MT que se han propuesto, pero antes de ello introduciremos un par de cuestiones que nos ayudarán a analizarlos mejor. La primera cuestión se refiere a la distinción que realizaron Churchland y Sejnowski (1992) entre *occurrent representations* (representaciones actuales o en uso) y *stored representations* (representaciones almacenadas). Churchland y Sejnowski consideran que las representaciones *occurrent* se hallarían en forma de patrones de activación neuronal sostenida mientras que las representaciones *stored* se hallarían en forma de configuraciones sinápticas entre neuronas (ver también, Davelaar, Goshen-Gottstein, Ashkenazi, Haarman y Usher, 2005; Ruchkin, Grafman, Cameron y Berndt, 2003). Los modelos que pretenden explicar los mecanismos neurofisiológicos subyacentes a la MT generalmente proponen que la MT consiste precisamente en una forma sostenida de activación neuronal (Davelaar *et al.*, 2005; Fuster, 1995; Goldman-Rakic, 1987; Smith y Jonides, 1999). Es decir, las representaciones contenidas en la MT serían representaciones *occurrent* en forma de patrones de activación neuronal sostenida.

Esta cuestión es importante a la hora de considerar el *priming*, ya que para que un determinado concepto pueda “activar” la representación de otro concepto, partimos del presupuesto de que el primero (el *prime*) ha de estar en forma *occurrent*. Un concepto en forma *stored* no podría enviar activación ni crear un atractor (en el caso de los modelos conexionistas de atractores) y por tanto generar un efecto de *priming*. También lo es en el caso del acceso léxico, ya que para que se pueda seleccionar el sentido adecuado de una palabra y acceder a él, la información que contribuye a la selección (o desambiguación) habría de estar en forma *occurrent*. Otro tanto habríamos de decir en el caso de la desambiguación sintáctica.

La capacidad de la MT

Una segunda cuestión, en la que nos extenderemos un poco, se refiere a la capacidad de la MT. Aunque existe un amplio consenso en aceptar que la capacidad de la MT está severamente limitada (Cowan, 2005), uno de los aspectos que más se ha debatido es su capacidad exacta de almacenamiento (para una revisión sobre las diferentes estimaciones de la capacidad de la MT, ver Cowan, 2005). Parece ser que el motivo principal por el que no hay acuerdo sobre la capacidad exacta de la MT es la intervención de procesos estratégicos o de “trucos mentales” en las tareas de recuerdo a corto plazo que permiten au-

mentar la cantidad de elementos (*chunks*) recordados. Según Cowan (2005) cuando se excluyen las situaciones experimentales que no garantizan la ausencia de estos procesos estratégicos, los datos empíricos sugieren que la capacidad “pura” de la MT sería en promedio de unos cuatro elementos. Aunque la capacidad de la MT se ha estudiado mayoritariamente con estímulos verbales, este mismo límite de cuatro elementos u objetos se ha encontrado también en los estudios que miden la capacidad de la MT puramente visual (ver Luck y Vogel, 1997).

Una pregunta de interés teórico para la que todavía no parece haber una respuesta definitiva es por qué existe este límite en la capacidad de la MT. En esta sección describiremos brevemente las explicaciones de tipo neurofisiológico que se han dado recientemente. Se ha argumentado (Lisman e Idiart, 1995; Luck y Vogel, 1997) que la razón del límite en la cantidad de objetos que pueden almacenarse en la MT de forma simultánea estaría en la necesidad de evitar interferencias debido a la forma que tiene el cerebro de representar los objetos. En el cerebro los objetos se representan de forma altamente distribuida debido a la especialización funcional de las diferentes áreas cerebrales (Felleman y Van Essen, 1991; Zeki, 1978). El caso más extremo estaría en el sistema visual de los primates. Las neuronas en cada una de las diferentes áreas visuales se han especializado, en cierto grado, en procesar solamente un determinado rasgo (color, forma, tamaño, orientación, movimiento, distancia, profundidad, etc.). Existe también abundante evidencia que indica que áreas cerebrales separadas participan en el reconocimiento y en la localización espacial de un mismo objeto más allá del córtex visual primario (Desimone y Ungerleider, 1989; Van Essen *et al.*, 1992). La correspondencia entre el tipo de objeto y su posición sólo es explícita en el nivel retiniano, ya que en los niveles superiores del sistema visual se pierde esta correspondencia y la información se segrega, como sucede en las vías del “qué” y del “dónde” que se localizan en el córtex temporal y parietal respectivamente (Goodale y Milner, 1992).

Este tipo de representaciones “distribuidas” presentaría un problema: ¿Cómo se puede obtener una representación unitaria de un objeto si sus diferentes rasgos se hallan representados de forma independiente y separada en el cerebro?: por ejemplo, si se está percibiendo un círculo verde pequeño y un triángulo rojo grande, cómo puede el cerebro “saber” que los rasgos “grande” y “rojo” y la forma “triángulo” (representados independientemente en áreas cerebrales distintas) forman parte del mismo objeto y que los rasgos “verde” y “pequeño” no forman parte de la representación de este objeto. A este problema se le conoce como el problema del *binding* (en ocasiones traducido al castellano como “problema del ligamiento”). En el cerebro se han de “ligar” los diferentes rasgos de alguna forma para generar la representación unitaria del objeto.

Diversos investigadores (Gray y Singer, 1989; von der Marlsburg, 1981) han propuesto una solución dinámica a este problema basada en el carácter temporal de la actividad neuronal. La idea básica es que las neuronas que representan los diferentes rasgos de un mismo objeto disparan sus potenciales de acción en sincronía (oscilan en una misma fase) pero de forma desincronizada (fuera de fase) respecto a las neuronas que representan los rasgos de otro obje-

to diferente (ver figura 1). Si no se resuelve correctamente este problema de *binding*, cuando hay varios objetos simultáneamente los rasgos de un objeto pueden pasar a ser percibidos como pertenecientes a otro objeto, fenómeno que se ha denominado *crossstalk*, conjunciones ilusorias o “catástrofe de superposición” (von der Malsburg, 1981).

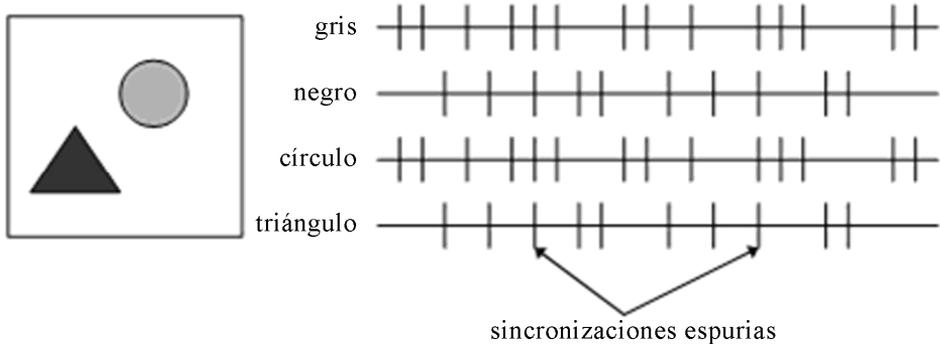


Figura 1. El disparo sincronizado de las neuronas se utiliza para codificar diferentes combinaciones de rasgos en la MT. Se muestran las respuestas de cuatro neuronas individuales que son selectivas al color o a la forma y que codifican los rasgos “gris”, “negro”, “círculo” y “triángulo” respectivamente. Cada una de las barras verticales representa una respuesta (un potencial de acción) de la neurona y el eje horizontal representa el tiempo. Las neuronas que codifican “gris” y “círculo” están ligadas conjuntamente ya que ambas se disparan en sincronía, al igual que sucede con las neuronas que codifican “negro” y “triángulo”. Puntualmente, sin embargo, ambos conjuntos de neuronas se activan o se disparan al mismo tiempo debido a una sincronización espuria, produciéndose una interferencia en la representación de los objetos.

Existe evidencia a favor de este mecanismo de *binding* en el sistema nervioso tanto en los sistemas sensoriales y motores (Gray y Singer, 1989; König y Engel, 1995) como en procesos de alto nivel que dependen del córtex frontal (Desmedt y Tomberg, 1994; Vaadia, Haalman, Abeles, Bergman, Prut, Slovin y Aertsen, 1995).

Podemos ahora entender la propuesta que comentábamos anteriormente: ¿Por qué la forma de representar objetos que tiene el cerebro limitaría la capacidad de la MT? Diferentes investigadores han argumentado desde un punto de vista neurofisiológico (Lisman e Idiart, 1995; Luck y Vogel, 1997) y computacional (Hummel y Biederman, 1992; Hummel y Holyoak, 2003; Hummel y Stankiewicz, 1998) que este límite evitaría que se produjese interferencia entre las señales de grupos de neuronas que representan rasgos de objetos diferentes y que tienen que mantenerse oscilando de forma desincronizada para poder diferenciarse. Según Luck y Vogel (1997), por ejemplo, si se almacenan demasiados objetos en la MT aumentará la probabilidad de que las neuronas que re-

presentan un rasgo de un objeto disparen en sincronía con las neuronas que representan un rasgo de un objeto diferente, debido a sincronizaciones espurias¹ (ver figura 1), dando lugar a una interferencia o a un solapamiento en las representaciones de objetos diferentes. Si se producen estas sincronizaciones espurias, el problema del *binding* no queda resuelto correctamente. No se garantiza que se puedan distinguir adecuadamente los rasgos que pertenecen a cada objeto diferente (qué va con qué) y pueden acabar produciéndose conjunciones ilusorias.

Un hecho a destacar es que este proceso de *binding* dinámico requiere atención (Hummel y Biederman, 1992; Hummel y Holyoak, 2003; Hummel y Stankiewicz, 1998; Treisman y Schmidt, 1982). Existe evidencia empírica que muestra que este mecanismo de *binding* puede establecerse de forma exitosa solamente si los elementos adecuados están en foco de atención, evitando que se produzcan conjunciones ilusorias (Cohen e Ivry, 1989; Treisman y Schmidt, 1982). Debido a que hay un número limitado de grupos de neuronas que pueden activarse al mismo tiempo y a la vez estar oscilando de forma desincronizada entre sí, sin que se produzca interferencia o sincronizaciones espurias o accidentales, la atención serviría para “decidir” los grupos de neuronas que han de participar en el proceso de *binding* dinámico (Hummel y Stankiewicz, 1998).

El mecanismo de sincronía de oscilación para resolver el problema del *binding* dinámico ha sido cuestionado (ver v.gr., Edelman e Intrator, 2003 para una propuesta alternativa). En cualquier caso, el problema del *binding* es una restricción importante que se ha de tener en cuenta para poder entender cómo se pueden estructurar las representaciones, por ejemplo a la hora de estudiar el fenómeno del *priming*, del acceso léxico y la desambiguación sintáctica. Supondremos que para que el efecto de *priming* pueda producirse, los rasgos del concepto que actúa como *prime* han de estar “ligados” en una representación unificada. Por otra parte el número de elementos en estado *occurrent* es muy reducido, lo que pondría en cuestión muchos modelos actuales: por ejemplo, los modelos de redes semánticas proponen que la activación se propaga de nodo a nodo, activando cada nodo todos los nodos con los que está conectado. Esto implicaría que en sólo dos pasos podemos llegar a tener activados muchísimos más elementos que sólo los 4 que se ha propuesto pueden estar en estado *occurrent*. Suponiendo que cada nodo está conectado con otros 5 (una estimación muy a la baja en el caso del cerebro humano) en dos pasos tendríamos 25 conceptos activados y en tres pasos tendríamos 125 conceptos activados. No parecería, pues, plausible un mecanismo de propagación de la activación (*spreading activation*) como el que proponen estos modelos y que se supone que en términos neurofisiológicos consistiría en propagación de activación a lo largo de circuitos neuronales (estado *occurrent*).

1. Hummel (ver v.gr., Hummel y Biederman, 1992; Hummel y Stankiewicz, 1998) se refiere a estas sincronizaciones espurias como “accidentales”.

La arquitectura de la MT

En el modelo de “múltiples componentes” de Baddeley (1986), que ha sido considerado como el modelo estándar de la MT durante mucho tiempo, el almacenamiento a corto plazo se produce en módulos especializados que están separados de los módulos de almacenamiento responsables de la memoria a largo plazo (MLP). Los módulos de almacenamiento a corto plazo tendrían un sustrato neurofisiológico diferente del de las estructuras responsables del almacenamiento a largo plazo. Desde esta concepción tradicional de la memoria, en la que se asumen módulos de almacenamiento separados, se ha utilizado la metáfora computacional de la “copia” para explicar cómo se mueve la información de un tipo de almacén a otro: por ejemplo, consideremos una oración como la de (1):

(1) El hombre ama a la mujer

En esta oración, el proceso de acceso léxico comportará acceder a la información (fonológica, léxica, sintáctica, semántica) asociada a *hombre*, que está almacenada en la MLP, y luego copiarla en un *buffer* de la MT para ser procesada momentáneamente y poder integrarla con otras palabras de la misma oración. La arquitectura funcional de estos modelos sería similar a la de un ordenador en el que el procesamiento está centralizado (v.gr., el ejecutivo central en el modelo de Baddeley) y la MLP es un elemento esencialmente pasivo. La información que el procesador central necesita manipular en cualquier momento se mantendría sólo activa en una memoria tipo RAM (v.gr., el bucle fonológico o la agenda viso-espacial).

Este tipo de arquitectura de la memoria recibe apoyo de los resultados obtenidos en experimentos psicológicos (efecto de primacía y recencia, efectos de interferencia por similitud fonológica, efecto de supresión articulatoria) y sobre todo se basa en datos neuropsicológicos (doble disociación entre MT y MLP). Se ha supuesto que la existencia de almacenes separados podría explicar adecuadamente estos resultados. Sin embargo, en los últimos 10 años se ha venido cuestionado esta interpretación, tanto de los datos experimentales como de los neuropsicológicos, y se han proporcionado interpretaciones alternativas en base a los modelos de MT en los que no se distinguen almacenes separados (v.gr., Brown, Preece y Hulme, 2000; Cowan, 2005; Davelaar *et al.*, 2005; Gupta, Lipinski y Aktunc, 2005; Oberauer, 2002; Ruchkin *et al.*, 2003). Por otra parte, existen numerosos estudios realizados con una serie de paradigmas, que se conocen en conjunto como “paradigmas de doble tarea” o “tareas complejas de MT” (ver Daneman y Carpenter, 1980), que ponen en serios aprietos a un modelo de MT como el que propone Baddeley (1986). La idea básica es saturar el componente (o en términos más generales los recursos) de almacenamiento de la MT haciendo memorizar, en una tarea secundaria, cierta cantidad de información (v.gr., reteniendo una lista de dígitos) mientras de forma simultánea se tiene que realizar otra tarea cognitiva, la tarea principal, que requiere computación o procesamiento (v.gr., aritmética mental). En el

modelo de Baddeley, el componente de la MT responsable de llevar a cabo la computación en este tipo de tareas es el ejecutivo central, que se caracteriza como un conjunto limitado de recursos generales de procesamiento (vendría a ser como el procesador de un ordenador). Como el ejecutivo central no tiene capacidad de almacenamiento por sí mismo, tiene que utilizar los almacenes “esclavos” (la agenda viso-espacial y el bucle fonológico) para almacenar la información necesaria cuando lleva a cabo una computación. Lo que predice el modelo que sucederá en un paradigma de doble tarea es que a medida que se vayan saturando los almacenes “esclavos”, con la tarea secundaria de almacenamiento, no quedará espacio para almacenar la información que necesita el ejecutivo central para llevar a cabo la tarea principal y que, por tanto, el rendimiento en la tarea principal se verá afectado.

Existen abundantes resultados empíricos, sin embargo, que muestran que la ejecución de la tarea principal no se ve afectada por la práctica saturación de los recursos de almacenamiento que conlleva la realización de la segunda tarea; recursos que se necesitarían para la realización de la primera tarea (Foos y Wright, 1992; Klapp, Marshburn y Lester, 1983; Oberauer, 2002; Oberauer, Demmrich, Mayr y Kliegl, 2001). Este hecho lleva a la siguiente paradoja: ¿Cómo el rendimiento en una tarea que parece claro que requiere almacenar información en la MT no se ve afectado cuando los recursos de almacenamiento de la MT están saturados por la realización de una tarea concurrente?

La alternativa a esta concepción tradicional de la MT la representa el modelo de Cowan (2005) y otros inspirados en él y de características similares (ver Ericsson y Kintsch, 1995; Garavan, 1998; Oberauer, 2002). En el modelo de Cowan (*embedded processes model*) la MT tendría dos componentes: (1) la parte activa de la MLP o conjunto de elementos que estarían activados y (2) un subconjunto más pequeño de esta parte activa que estaría en el foco de atención² (ver figura 2). Se asume que sólo el foco de atención tiene capacidad limitada. La capacidad del foco de atención sería según Cowan (2005) en promedio de unos cuatro elementos. No habría un límite de capacidad en cuanto al número de elementos de MLP que pueden estar “activos”, aunque estas representaciones dejarían de estar activas debido a un proceso de decaimiento de la activación o de interferencia (para una propuesta similar ver Anderson *et al.*, 1996).

Hasta ahora habíamos distinguido entre dos tipos de representaciones: las *occurrent* y las *stored*, pero en el modelo de Cowan se necesita introducir un tercer tipo de representaciones ya que dentro de la MT Cowan distingue dos tipos de representaciones, las que están activas pero no en el foco de atención y las que están activas y se hallan en el foco de atención. Parece claro que las representaciones en formato *occurrent* (en forma de patrones de activación sostenida) serían las que están en lo que Cowan llama foco de atención. Lo que Cowan

2. El “foco de atención” en el modelo de Cowan se correspondería con la “región de acceso directo” en el modelo de Oberauer (2002), ya que Oberauer establece otra distinción más dentro de la MT y considera que habría un foco de atención separado de capacidad más limitada todavía. Sin embargo, como el propio Oberauer señala, esta diferencia entre ambos modelos no es significativa a la hora de explicar el límite en la capacidad de la MT, ni tampoco para explicar cómo se llevaría a cabo una tarea compleja de MT.

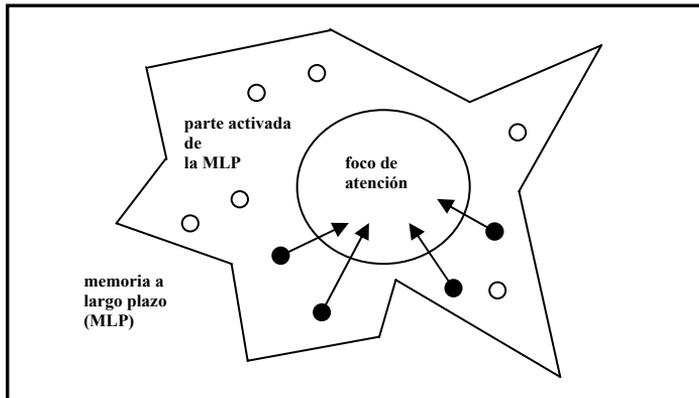


Figura 2. Ilustración del modelo concéntrico o anidado de Cowan (1988, 1995, 1999, 2005) en la que se muestra cómo la información de la parte activada de la memoria a largo plazo (dentro de la figura irregular dentada) es llevada hasta el almacén de capacidad limitada, el foco de atención (círculo grande). Los círculos pequeños de color negro con flechas son los únicos que pueden entrar en el foco de atención

denomina representaciones activas pero que no estarían en foco de atención también parece claro que no pueden estar en formato *occurrent*, ya que, según Cowan, no habría limitación para el número de representaciones que podrían encontrarse en este estado. Se entraría en contradicción con los estudios neurofisiológicos (la limitación en el número de objetos debido al problema de la *binding*) que hemos visto en la sección anterior si consideráramos que en estado *occurrent* pueden existir un número ilimitado de elementos. Vamos a llamar a este tipo de representaciones *stored** sin preocuparnos de momento por el mecanismo neurofisiológico requerido para su realización. La característica que distinguiría las representaciones en estado *stored** de las *stored* (el resto de representaciones de la MLP) sería la rapidez con la que se podría acceder a ellas. Tenemos así tres tipos de representaciones. El primer tipo serían las representaciones *occurrent* (en foco de atención, y en forma de patrones de activación sostenida). El número de elementos que podrían estar en forma *occurrent* sería en promedio de cuatro. En segundo lugar tendríamos las representaciones en forma *stored** que serían mucho más accesibles que las del tercer tipo que hemos denominado *stored*.

Según esta concepción concéntrica de la memoria, recuperar un elemento de la MT, ya sea para su recuerdo o su manipulación, equivale a colocar este elemento dentro del foco de atención. En términos neurofisiológicos, al recuperar un elemento de memoria pasamos de un estado *stored** o *stored* a uno *occurrent*. Que la capacidad del foco de atención sea aproximadamente de cuatro elementos y que en principio no haya límite para las representaciones en formato *stored** encaja con la idea de que el foco de atención corresponde a patrones de activación neuronal y con lo que dicen los modelos de MT basados en la sincronización de la

actividad neuronal que vimos en la sección anterior. También encaja con la idea de que el proceso de *binding* requiere atención. El número de representaciones en formato *occurrent* o en foco de atención está limitado porque hay un número limitado de patrones de activación que se pueden hacer oscilar en la misma fase y ser distinguidos al mismo tiempo (al estar en fases distintas) de otros patrones de activación sin que llegue a producirse interferencia o sincronizaciones espurias entre los patrones de activación que corresponden a diferentes objetos.

Debido a esta limitación en el número de objetos que pueden estar en un estado de máxima accesibilidad o en foco de atención, el éxito en la computación en una determinada tarea dependerá de la habilidad para controlar la atención y dirigirla hacia los objetos que son más relevantes para esta tarea. La función que juega la atención respecto a la memoria sería por tanto similar a la que tiene en percepción, actuando como una especie de “portero” que determina qué objetos pueden acceder a la MT y ocupar el espacio severamente limitado del foco de atención (Awh, Vogel y Oh, 2006). Esta arquitectura alternativa de la MT, por otra parte, se ajusta de forma natural a la caracterización que hacen algunos investigadores de la MT como “control atencional” (Garavan, 1998; Kane y Engle, 2003). Según esta concepción, la interacción entre las funciones de almacenamiento y computación de la MT está mediada por la atención selectiva al conjunto de objetos o elementos en la MT que deben ser manipulados para llevar a cabo la computación. Si tenemos en cuenta la caracterización de la MT como “control atencional” y la función de “portero” que desempeña la atención respecto a la MLP, que pueden considerarse como dos caras de una misma moneda, la paradoja que se planteaba al interpretar los resultados obtenidos mediante el denominado paradigma de la doble tarea queda resuelta. Si se asume un modelo de la MT como el que proponen Cowan o Oberauer, las dos tareas que se realizan simultáneamente no comparten un recurso de almacenamiento limitado, ya que en estado *stored** se podría almacenar una cantidad importante de elementos. En un paradigma de doble tarea, el rendimiento en la tarea principal se verá afectado sólo si la otra tarea requiere manipular información en el foco de atención, es decir, cuando ambas tareas requieren manipular representaciones en formato *occurrent*. Esto ocurriría, por ejemplo, cuando en la tarea secundaria los dígitos que deben ser recordados tienen que actualizarse mediante el resultado de las operaciones aritméticas que se están computando. La tarea secundaria no afecta al desarrollo de la tarea principal si la información que tiene que recordarse no necesita ser manipulada en el foco de atención y se mantiene en estado *stored** (ver también, Anderson, Reder y Lebiere, 1996; Awh, Vogel y Oh, 2006; Oberauer, 2002; Oberauer *et al.*, 2001).

Por último, desde un punto de vista neurofisiológico, los modelos de memoria, como el de Baddeley, presentan dos dificultades principales. En primer lugar, estos modelos proponen una distinción estructural, típica en la arquitectura computacional clásica, entre las funciones de memoria y de procesamiento. Pero esta distinción no se corresponde bien con la arquitectura del cerebro, en la que estas funciones parece que están altamente distribuidas y son llevadas a cabo por el mismo sustrato neuronal (Churchland y Sejnowski, 1992; Damasio, 1989; Rumelhart y McClelland, 1986). En segundo lugar, no resulta claro

cómo se realizaría la metáfora de la copia de información entre las diferentes estructuras de almacenamiento.

Una parte importante de los estudios neurofisiológicos han utilizado de forma casi exclusiva el modelo de múltiples componentes de Baddeley para entender las bases neuronales de la MT (Henson, 2001; Smith y Jonides, 1999). Generalmente se han situado los *buffers* para el almacenamiento de información en las zonas posteriores del cerebro (v.gr., córtex parietal y temporal) mientras que las operaciones llevadas a cabo por el ejecutivo central se han atribuido al córtex prefrontal (D'Esposito *et al.*, 2000; Smith y Jonides, 1999). Pero mientras existe un consenso generalizado en aceptar que el córtex prefrontal constituiría una parte importante de la MT (D'Esposito *et al.*, 2000; Fuster, 1995; Goldman-Rakic, 1987; Smith y Jonides, 1999), no está nada claro cómo se realizaría en el nivel cerebral el modelo de estructura modular que propone Baddeley (Chein, Ravizza y Fiez, 2003; Pasternak y Greenlee, 2005).

Un modelo como el de Cowan o el de Oberauer, en los que la MT constituye la parte activada de la MLP, puede resolver de forma natural muchas de las dificultades anteriormente planteadas y, además de ser más integrador, es también plausible neurofisiológicamente. Según Cowan (2005), las áreas del córtex prefrontal estarían relacionadas con el control atencional, estableciendo punteros a la información que se encuentra representada en las áreas posteriores del cerebro. Estos punteros señalarían aquella información que tiene que pasar de un estado *stored** o *stored* a estar en formato *occurrent*. De esta forma se recrearían los conceptos o los objetos que se han de colocar en foco de atención. La MT constituiría esencialmente atención que se focaliza en representaciones internas, es decir, en las representaciones que se hallan almacenadas en la MLP (Fuster, 1995). El problema de la copia de información queda resuelto y no se han de proponer estructuras neurofisiológicas diferenciadas para las representaciones sensoriales y para las representaciones activas en MT (ver también, Damasio, 1989; Fuster, 1995; Ruchkin *et al.*, 2003; Pasternak y Greenlee, 2005). Existe evidencia empírica que apoyaría una arquitectura de la MT como la que propone Cowan (ver v.gr., Chein *et al.*, 2003; Kane y Engle, 2003; Ruchkin *et al.*, 2003). El modelo de Cowan no distingue entre estructuras de almacenamiento diferentes y considera que la atención actúa como portero introduciendo determinados elementos en el foco de atención de la MT (Awh, Vogel y Oh, 2006). Esta propuesta es coherente con la propuesta que hizo Damasio (1989) sobre cómo el cerebro integra la información que está en áreas separadas y resuelve de forma más general el problema del *binding*. La "reconstrucción" de un objeto o de un concepto (ver Figura 3) en la MT tendría lugar en las mismas áreas sensoriales (o motoras) encargadas de su percepción y por tanto se evita la "copia" de información de un sustrato neuronal encargado de la codificación sensorial a otro responsable de su almacenamiento en MT. Damasio también propone la misma solución que hemos visto anteriormente basada en la sincronización de disparo para resolver el problema del *binding*. La representación unitaria de un objeto cuando se utilizan representaciones distribuidas para codificar sus diferentes rasgos se puede conseguir haciendo que las neuronas que representan estos rasgos oscilen en

una misma fase. Según Damasio, las áreas cerebrales anteriores (v.gr. córtex prefrontal) estarían implicadas en la integración de la actividad de las diferentes neuronas o de los grupos de neuronas que codifican los diferentes rasgos de un objeto, pero esta integración no se lograría mediante áreas “finales” de integración que acumularían toda la información que se encuentra segregada en las áreas posteriores del cerebro. Damasio señala la importancia de la bidireccionalidad en las conexiones que se establecen entre el córtex anterior y el posterior, y la existencia de numerosas conexiones tipo *feedback* sería consistente con la función del córtex prefrontal como un sistema de punteros que se encargarían de integrar los diferentes rasgos de un mismo objeto que se encuentran separados en el córtex posterior y en otras áreas.

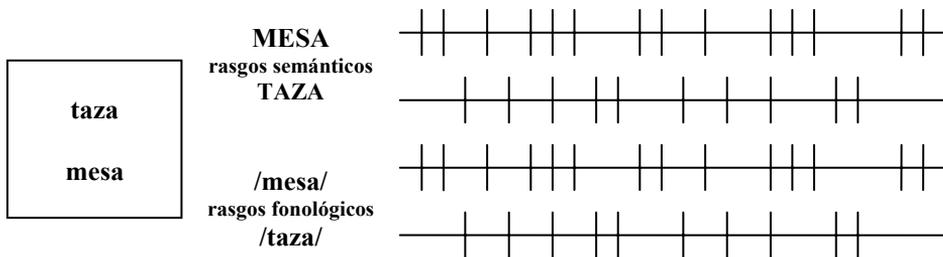


Figura 3. Al igual que sucede con los diferentes rasgos de un objeto, cuando reconocemos una palabra los diferentes rasgos de ésta (fonológicos, sintácticos, semánticos) se han de ligar en la MT mediante un mecanismo de *binding*. Los diferentes rasgos semánticos de la palabra *mesa*, correspondientes al concepto de MESA, tendrían que ligarse con los rasgos fonológicos correspondientes (/mesa/).

En la siguiente sección analizaremos cómo el problema de la composicionalidad se puede resolver mediante una arquitectura de la MT como la que hemos descrito, en fenómenos tan diversos como la aritmética mental, el razonamiento, el reconocimiento visual de objetos, el procesamiento sintáctico y el proceso de desambiguación del sentido. Esto nos permitirá introducir más restricciones que tendríamos que considerar a la hora de elaborar un modelo sobre el *priming* semántico, el acceso léxico o la desambiguación sintáctica.

Composicionalidad

El término “composicionalidad” fue formulado por Frege con el ánimo de proporcionar una manera formal de obtener el significado de expresiones matemáticas. En su obra *Foundations of Arithmetics*, Frege (1903) establece también un paralelismo entre la forma en que se puede obtener el significado de formulas matemáticas y la forma en la que se podría obtener el significado de expresiones lingüísticas.

En el caso de la aritmética, el concepto de composicionalidad es fácil de comprender. Consideremos una expresión como la que se muestra en (2c). Cualquier sistema (natural o artificial) que tenga conocimientos básicos de aritmética puede resolver esta expresión aunque no la haya visto nunca antes. Esto es debido a que la aritmética es composicional. Si un sistema es capaz de resolver las expresiones simples que se muestran en (2a) y (2b) también debería ser capaz de resolver las mismas expresiones simples cuando forman parte de una expresión más compleja, como (2c), aunque jamás haya visto esta expresión antes.

- (2) a. $(2 + 3)$
 b. $(8 - 2)$
 c. $7 + ((8 (2 + 3) + (9 - 6)) (8 - 2))$
 d. $4 + (3 ((8 - 2) + 7) - (2 + 3))$

Lo que el principio de composicionalidad dice es que expresiones simples o básicas como las mostradas en (2a) o (2b) se reconocen como tales y se resuelven de forma invariante independientemente del contexto en el que se hallen. Gracias a esta propiedad es posible resolver (o generar) cualquier expresión aritmética, por compleja que ésta sea, conociendo cómo se resuelven o generan las expresiones básicas que la constituyen y las reglas que las combinan. El número posible de expresiones aritméticas es infinito y todas ellas se pueden generar a partir de la combinación de un número finito de expresiones simples y de reglas de combinación. En lo que sigue vamos a considerar cómo una arquitectura de la MT como la que hemos descrito permitiría aplicar el principio de composicionalidad de una forma simple y sencilla con plausibilidad psicológica. Consideremos la expresión en (3). Hemos representado en **negrita** lo que, según se propone, estaría en el foco de atención, mientras que lo que está *background* (en estado *stored**) se ha representado con caracteres normales. Una forma simple de resolver la expresión en (3) consistiría en focalizar la atención en el primer término ($3+4$), como se muestra en (3a), resolverla y dejar el resultado en *background*. Luego focalizar la atención en el segundo término, como se muestra en (3c), resolverlo y almacenar de nuevo el resultado. Finalmente, focalizar la atención en los dos resultados y operarlos.

- (3) a. $(\mathbf{3+4}) (7-9)$
 ↓
 b. $7 (7-9)$
 c. $7 (\mathbf{7-9})$
 ↓
 d. 7×2
 e. $7 \times 2 \rightarrow 14$

Al hacerlo de esta forma tratamos cada expresión de forma independiente del contexto en el que se halla y luego combinamos los resultados parciales siguiendo el principio de composicionalidad. Esta forma de resolverlo es pre-

cisamente para la que Oberauer (como vimos en el apartado referente a la arquitectura de la MT) ha encontrado abundante evidencia empírica. Lo que Oberauer muestra en sus experimentos es que en “foco de atención”, en estado *occurrent*, se encuentra en cada momento solamente un operador y los dos argumentos de este operador que se van a computar. El resto está en *background* o estado *stored**. Ésta es la forma en que el cerebro aplica el principio de composicionalidad. Teniendo una expresión básica en foco de atención y nada más, la podemos tratar de forma independiente e invariante al contexto en que se halla. Si tuviéramos otros elementos de otras expresiones vecinas en estado *occurrent* o en foco de atención, la tarea se dificultaría enormemente o sería imposible llevarla a cabo. Esto aparece de forma más clara si intentamos hacer un modelo computacional plausible que lleve a cabo el cómputo. En el *input* a este modelo solamente podemos colocar los argumentos y el operador, no más información.

La aritmética seguiría también los principios de productividad y sistematicidad propuestos por Fodor y Phylyshyn (1988). De hecho estos dos términos derivan de forma natural del principio de composicionalidad propuesto por Frege y en adelante hablaremos solamente de composicionalidad.

Se considera que muchos aspectos de la cognición son composicionales (la visión, el pensamiento, la actividad motora, etc.), y no solamente el lenguaje o más específicamente la sintaxis como considera Chomsky (Bienenstock, Geman y Potter, 1997; Harrison, 2005).

La composicionalidad, en el caso del lenguaje, presenta más problemas que en el caso de la aritmética. El mismo Frege propuso otro principio, el principio de la contextualidad (“nunca preguntes por el significado de una palabra aislada”), que parece contradecir el principio de composicionalidad. Muchos aspectos del significado de una frase (por ejemplo, el sarcasmo o la ironía) no se pueden obtener sobre la base de combinar el significado de las palabras aisladas que la forman. Por otra parte, el lenguaje es muy ambiguo, la mayoría de las palabras tienen muchos sentidos y solamente en contextos adecuados quedan desambiguadas. Pero no cabe duda de que si queremos explicar la capacidad de comprender frases nuevas, frases que pueden ser arbitrariamente complejas, hemos de tener en cuenta alguna forma de composicionalidad. En un enfoque cognitivo, los dos principios serían perfectamente compatibles, si suponemos que el proceso composicional de formación del significado, a partir de combinar los significados de las palabras, se puede completar con procesos de desambiguación del sentido y procesos de inferencia que añaden nueva información o modifican el significado obtenido por mecanismos puramente composicionales.

Para considerar el lenguaje composicional deberíamos poder definir, al igual que en el caso de la aritmética, expresiones o unidades básicas y comprobar que pueden resolverse, al menos inicialmente, de forma independiente del contexto en el que se hallan. Estas expresiones básicas serían como los bloques de construcción (*building blocks*) a partir de los cuales, y combinándolos adecuadamente, podríamos, como en el caso de la aritmética, construir un número prácticamente infinito de oraciones que podrían ser todo lo arbitra-

riamente complejas que deseáramos. Una de las cuestiones que tendría que definir una teoría adecuada del lenguaje sería la naturaleza de estas expresiones básicas. En una teoría como la chomskiana, las unidades básicas son los constituyentes o sintagmas. Sin embargo, otras teorías como la *Lexicalized Tree Adjoining Grammar*, LTAG, (Joshi, 1999), la *Role and Reference Grammar*, RRG, (van Valin y LaPolla, 1997) o las *Construction Grammars* (Fillmore, 1988; Goldberg, 1995) definen estructuras más complejas como unidades básicas, según veremos más adelante. Los criterios a seguir para definir estas unidades básicas vendrían en parte determinados por cómo pueden “encapsular” y especificar dependencias de diverso tipo (sintácticas, semánticas, morfosintácticas...). No podemos extendernos mucho en este tema. Solamente apuntaremos algunas cuestiones ilustrándolas con ejemplos sencillos. Consideremos la frase en (4). En esta frase aparece una de las ambigüedades que a un sistema artificial de procesamiento de lenguaje le resultan más difíciles de tratar y que más atención ha recibido en la lingüística computacional actual. Se la denomina *PP attachment ambiguity* o ambigüedad en la adjunción del sintagma preposicional. En (4a) el sintagma preposicional *with a fork* se adjunta al verbo, ya que modifica el modo en que se come. En el caso de (4b) el sintagma preposicional se adjunta al sintagma nominal *the salad* ya que especifica cómo es la ensalada. Utilizando información de subcategorización y/o semántica y/o de “conocimiento del mundo”, estas ambigüedades puede resolverse adecuadamente.

- (4) a. *Mary ate the salad with a fork*
 b. *Mary ate the salad with croutons*

Uno de los problemas de las oraciones reales, por ejemplo las que nos podemos encontrar en los periódicos, es que no son tan simples como las mostradas en (4). Consideremos las oraciones en (5).

- (5) a. *The company requested a treatment for watches.*
 b. ***The company*** *that is seeking to borrow money from another company had requested duty-free treatment for many types of watches.*
 c. ***The company*** *that is seeking to borrow money from another company had requested during a December Science Steering Group meeting duty-free treatment for many types of watches.*

Parece claro que en (5a), el sintagma preposicional *for watches* se ha de adjuntar al nombre predicativo *treatment* y no al verbo *requested*. También parece claro que con estos cuatro elementos: el verbo (*requested*, en este caso), el primer nombre (*treatment*), la preposición (*for*) y el segundo nombre (*watches*) se tiene suficiente información para decidir la adjunción del sintagma preposicional. Parecería claro que éste es también el caso en las oraciones en (5b) y (5c). Hindle y Roth (1993) idearon un método estadístico para determinar la adjunción de este tipo de sintagmas preposicionales a base de computar las co-ocurrencias de estos 4 elementos. En frases más complejas el “truco” consiste en extraer precisamente estos cuatro elementos e ignorar el resto de

elementos de la oración. En los ejemplos en (5b) o (5c) estos 4 elementos están en negrita. Es lo único que se extrae (o se le “presta atención”) para contar las co-ocurrencias ignorando todo lo demás. Este “truco” permitió obtener un sistema artificial, y por primera vez, unos resultados sorprendentemente buenos en esta tarea y es el “truco” que se sigue todavía utilizando para desambiguar el sintagma preposicional (Collins y Brooks, 1995; Alegre, Sopena y Lloberas, 1999). Pero ¿se trata simplemente de un truco, como ha sido considerado, o se nos está revelando alguna cosa interesante sobre el procesamiento del lenguaje? Parece claro que los elementos de la oración de relativo como *seeking*, *borrow* o *money* no contribuyen en nada a la decisión sobre dónde adjuntar el sintagma preposicional; es información *irrelevante*. Lo mismo podemos decir de los otros modificadores. Parece que todo se resuelve en un nivel local, “prestando atención” únicamente a los 4 elementos que pueden potencialmente formar una proposición y cuyo predicado sería el verbo. La proposición se muestra en (6). Lo que el sistema decide es si *for watches* puede actuar como ARG3 del verbo. Si la decisión es negativa, este argumento se borra de esta posición argumental. La adjunción de este sintagma preposicional se resolvería en esta “unidad” de 4-elementos ignorando el contexto en el que se halla inserta (por ejemplo, 5.a, 5.b o 5.c). Parece que podemos aplicar el principio de composicionalidad.

(6) PRED: *Request*; ARG1: *the company*; ARG2: *treatment*; ARG3: *for watches*

Consideremos ahora las frases en (7). Las frases y expresiones en (7a-d) pueden considerarse partes elementales de las oraciones complejas en (7e-h). La oración simple (7a) sería la oración principal (está en negrita) de estas oraciones complejas (7e-h). Parece que algún principio de composicionalidad tendría que aplicarse, ya que es razonable suponer que si un sistema puede “comprender” la frase en (7a) también debería comprender esta frase aunque esté insertada en estas oraciones nuevas más complejas (7e-h). De igual forma, si el sistema ha aprendido que el sentido correcto del verbo *take* en (7a) es “capturar”, también habría de saber que éste es el sentido correcto de *take* en las frases complejas (7e-h). La forma de entender este comportamiento se explicaría suponiendo que existe un principio de composicionalidad. El significado de una expresión básica se obtiene de forma independiente del contexto en el que se halla insertada. Ésta es una propiedad que no tienen los modelos conexionistas de procesamiento sintáctico diseñados hasta el momento. Estos sistemas pueden aprender a analizar y “comprender” una frase como (7a) pero serían incapaces de generalizar esta comprensión a la oración principal de frases nuevas como las mostradas en (7e-h). Podríamos incluso enseñarles a comprender todas las frases en (7) con excepción de, digamos, la (7g) y a pesar de ello serían incapaces de generalizar lo aprendido a esta frase nueva (7g). Este hecho se puede comprobar fácilmente llevando a cabo simulaciones con estos modelos. En ellas observaríamos de forma reiterada que no pueden generalizar lo que han aprendido, incluido lo que han aprendido en frases complejas como (7h), a una frase nueva como (7g). No podrían ni analizar bien su

oración principal. Una excepción la constituye el modelo de Miikkulainen (1996) que propone una solución parecida a la que estamos considerando. La composicionalidad, tal como la hemos definido arriba es, pues, una propiedad deseable, ya que no es razonable suponer que los hablantes de una lengua se comporten, como los modelos conexionistas, de una forma no composicional en ejemplos como los mostrados en (7).

- (7) a. ***The carpenter took a prisoner.***
 b. *The coast of northern Sumatra.*
 c. *The old carpenter.*
 d. *The young prisoner.*
 e. ***The carpenter who lived on the coast took a prisoner.***
 f. ***The old carpenter took a young prisoner.***
 g. ***The old carpenter who lived on the coast took a young prisoner.***
 h. ***The old carpenter who lived on the coast of northern Sumatra took a young prisoner.***

En las LTAG (Joshi, 1999), las unidades básicas o bloques de construcción que permitirían llevar a cabo el principio de composicionalidad son lo que se denomina “árboles elementales” (*elementary trees*). Existen dos tipos de árboles elementales, los iniciales y los auxiliares (que incorporan los modificadores). Como se muestra en Sopena y Lloberas (en preparación), cada uno de estos árboles elementales se puede hacer corresponder con una estructura predicado-argumento (una proposición). En el caso de la RRG (van Valin y LaPolla, 1997) ocurre, de forma parecida, que las unidades básicas que se proponen se pueden hacer corresponder aún de forma más clara con proposiciones simples (un predicado y sus argumentos). En la teoría que hemos desarrollado (Alegre 2004; Sopena y Alegre, 2000; Sopena y Lloberas, en preparación), las unidades básicas se definen también como aquellas expresiones que se corresponden con una sola proposición. Hemos visto que en el caso de la aritmética, las expresiones básicas estarían formadas por un operador y sus argumentos; en el caso del lenguaje, una expresión básica también se podría hacer corresponder con una proposición (un predicado y sus argumentos). Las expresiones que definen una proposición constituyen el dominio de localidad que permite resolver las dependencias sintácticas y semánticas relevantes.

Al igual que hemos visto en el caso de la aritmética, podemos implementar un principio de composicionalidad para el lenguaje que sea psicológica y neurofisiológicamente plausible utilizando el modelo de MT que hemos descrito. Consideremos las frases en (8). En (8a-k) se muestran qué elementos predice la teoría estarían en foco de atención a medida que se va procesando la oración. En el foco de atención tendríamos que tener solamente las palabras que se corresponden con una unidad básica para poder aplicar el principio de composicionalidad. En (8a-k), la palabra subrayada sería la palabra que actualmente se está leyendo. Los contenidos del foco de atención (estado *occurrent*) constituirían el dominio de localidad en el que las dependencias semánticas y sintácticas se resuelven. Como se puede apreciar en (8b) tenemos en foco de atención solamente una proposición formada por un predicado (el

adjetivo *tall*) y su argumento (*carpenter*). En (8c) esta proposición ha desaparecido del foco de atención y el fragmento de la oración principal procesado hasta ahora se ha colocado de nuevo en foco de atención (***the carpenter***). Con el sintagma preposicional *with the handbag*, se abre una nueva proposición que se coloca en foco de atención (8d-e). La proposición principal ha pasado a *background* o al estado *stored**. Cuando aparece el verbo *gave* se coloca de nuevo en foco de atención la oración principal (***the carpenter gave***). En (8) no hemos representado todos los focos de atención ya que sería demasiado largo. Considerar la situación en (8k). La palabra que se está leyendo es *book* y lo que tenemos en foco de atención es la oración principal (***the carpenter gave the nurse the book***). Al igual que en el caso de la aritmética, cada expresión básica se procesa de forma independiente del contexto en el que se halla insertada. Las expresiones básicas vecinas son en general información *irrelevante* para el procesamiento sintáctico y semántico de una expresión básica dada y tenerla en cuenta haría intratable el procesamiento de esta expresión. Como se ve en los ejemplos, cada expresión básica se corresponde con una proposición: en foco de atención tenemos, en cada momento, una sola proposición (un predicado y sus argumentos) que se corresponde con una expresión básica. Es lo que permite la composicionalidad y lo que da capacidad al sistema para comprender o generar infinitas frases partiendo de medios finitos (una descripción más detallada se puede encontrar en Alegre 2004; Sopena y Alegre, 2000; Sopena y Lloberas, en preparación).

- (8) a. *The tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*
 b. *The tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*
 c. ***The** tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*
 d. *The tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*
 e. *The tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*
 f. ***The** tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*
 g. ***The** tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*

 j. ***The** tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*
 k. ***The** tall carpenter with a handbag gave the nurse who works at the hospital and lives in Elm Street the book.*

Una forma de probar que un principio de composicionalidad como el que hemos sugerido es plausible, es mostrar que un sistema artificial que lo utilice ha de ser muy eficiente con problemas reales, al igual que los seres humanos. Lo que nos dice este principio de composicionalidad es que las dependencias sintácticas

y semánticas se habrían de computar localmente en el seno de una unidad básica. En este dominio de localidad se hallaría la información *relevante* para resolver los problemas que plantea el análisis sintáctico y semántico. En términos psicológicos, esta idea de localidad se corresponde con el concepto de foco de atención. A este respecto es conveniente resaltar que los mejores resultados (con mucha diferencia) hasta ahora obtenidos por sistemas artificiales en tareas de análisis sintáctico con textos reales se han conseguido con esta idea de localidad (Collins, 1996; Charniak, 1997). También en el caso de tareas de desambiguación sintáctica o semántica (Alegre, 2004; Alegre *et al.*, 1999; Lloberas, en preparación; López-Moliner, 1998; Sopena, Díaz, Lloberas y Moliner, 1998). Como se describe en Sopena y Lloberas (en preparación), de nuevo, en estos modelos podemos asimilar la idea de localidad a la de foco de atención.

El problema del binding reconsiderado

Una dificultad sería que tienen que afrontar los diferentes modelos de MT, los que asumen una arquitectura tradicional como la de Baddeley y los modelos alternativos como el de Cowan, es que no ofrecen una explicación detallada, paso a paso, del procesamiento que tiene lugar en la MT cuando se tienen que integrar diferentes elementos, manteniéndolos en un estado de máxima accesibilidad, para computar una función compleja. La estructura de almacenamiento de la MT se ha de entender como un sistema activo de procesamiento, una pizarra donde se llevan a cabo computaciones, y no como un simple almacén donde se listan los diferentes elementos que van entrando. Ni Cowan, ni Baddeley, por ejemplo, tienen en cuenta el problema del *binding*. En el caso del lenguaje, por ejemplo, una teoría de la MT debería especificar los procesos en la MT que posibilitan el procesamiento de una oración. Se ha considerado (ver v.gr., Hummel y Holyoak, 2003) que la estructura en la MT vendría impuesta por el mecanismo de *binding* necesario para ligar los diferentes rasgos de los elementos que están en foco de atención. En el caso de tener una proposición en foco de atención, mediante la sincronización de disparo se consigue ligar las representaciones de los roles (variables) con las representaciones distribuidas de sus respectivos *fillers* (valores). Consideremos, por ejemplo, que en foco de atención tenemos la proposición correspondiente a la frase *el hombre ama a la mujer*. Esta proposición se halla representada en (9).

(9) PRED: ama; ARGUMENTO1: hombre; ARGUMENTO2: mujer

Mediante este mecanismo de *binding*, las unidades que representan los diferentes rasgos (sintácticos, semánticos, etc.) de “hombre” se dispararían sincronizadas en una misma fase, en una fase que podríamos denominar ARG1 (o rol agente). Las unidades para “mujer” también se dispararían en fase pero de forma desincronizada, en una fase distinta a la que se disparan las unidades para “hombre”. Lo mismo sucedería con las unidades que representan el predicado (*ama*). Las relaciones entre los roles y los *fillers* correspondientes de

una misma proposición se establecen de forma independiente, mediante las diferentes fases de oscilación. Por otra parte, como sólo los elementos de una única proposición pueden estar en estado *occurrent*, se evita la interferencia o el ruido con el resto de representaciones que no pertenecen a esa proposición.

Conclusiones

Hemos constatado en este estudio que muchos de los modelos que se han propuesto en las diferentes disciplinas que conforman la ciencia cognitiva, como la inteligencia artificial, la psicología o la neurociencia y en áreas de estudio tan distintas como el procesamiento visual, el razonamiento, el procesamiento del lenguaje o los estudios sobre la memoria, aparecen, aunque de forma independiente, ya que en general estos estudios se ignoran mutuamente, argumentos de todo tipo (computacionales y experimentales) que convergen todos en un misma dirección, al defender, al menos en alguno de sus aspectos, una estructura de la MT como la que hemos descrito.

Los argumentos son muy diversos y algunos han sido formulados en contextos muy alejados de la Psicología, pero muy valiosos también si se consideran desde un punto de vista lo suficientemente general. Los argumentos de tipo computacional se han originado en áreas en principio tan distintas como el razonamiento (Hummel y Holyoak, 2003; Shastri y Ajjanagadde, 1993), la desambiguación sintáctica (Hindle y Roth, 1993; Collins y Brooks, 1995; Sopena *et al.*, 1998; Alegre *et al.*, 1999), la desambiguación del sentido (López Moliner, 1998; Alegre, 2004) o el análisis sintáctico (Collins, 1996; Charniak, 1997; Sopena y Alegre, 2000; Sopena *et al.*, 1998; Sopena y Lloberas, en preparación, Miikkulainen, 1996). Desafortunadamente, por problemas de espacio, no podemos entrar en los detalles de esta argumentación, pero sí resaltar que las características que en estos modelos se propone que ha de tener la MT son coherentes con toda la evidencia neurofisiológica y psicológica comentada en las secciones anteriores y con lo comentado sobre el principio de composicionalidad.

Se ha especulado que la limitación en la capacidad de nuestra memoria de trabajo (o del foco de atención) no debería considerarse una anomalía de la que culpar a las características de nuestras neuronas o a nuestro tejido nervioso, sino que más bien esta limitación constituiría una solución ingeniosa para poder implementar un principio de composicionalidad. El cerebro, por ejemplo, de un cangrejo no estaría dotado de mecanismos de atención sofisticados y no podría fragmentar el mundo en unidades invariantes recombinables de formas casi infinitas para formar “objetos” más complejos. Parecería que la inteligencia podría surgir con esta capacidad de fragmentar, mediante la atención, el “mundo” en unidades básicas que pueden recombinarse. Esto mismo ocurriría en la visión para poder reconocer objetos de forma invariante en una diversidad de escenas visuales y poderlas estructurar, en el razonamiento, con el fin de establecer analogías o hacer inferencias y, en general, “crear pensamiento”, en el caso del lenguaje.

Los estudiosos de procesos como el acceso léxico, la desambiguación sintáctica y el *priming* semántico deberían tener en cuenta esta caracterización de

la MT para restringir el abanico de posibilidades a la hora de elaborar modelos sobre estos fenómenos. Parece claro que modelos de tipo interactivo como los propuestos por McDonald, Pearlmutter y Seidenberg (1994) no se ajustarían a estas restricciones. La cantidad de información que se propone en estos modelos debería estar en estado *occurrent* o en foco de atención supera con creces la limitación en la capacidad de la MT que hemos descrito en este trabajo. Igualmente ocurre con los modelos sobre el *priming* semántico, tanto en el caso de los modelos de atractores como en los modelos de redes semánticas. En el caso de los modelos de atractores, es curioso que en la caracterización de los rasgos que un concepto tiene se incluyan algunos tan complejos como “vive en el Polo”, “pone huevos” o “vive en el bosque”. Estos rasgos complejos superan de nuevo la capacidad de la MT. Rasgos como “los pingüinos ponen huevos” son una proposición completa y constituirían por sí mismos una unidad de atención que saturaría la capacidad de la MT. Hemos de pensar que los diferentes aspectos de un concepto se exploran serialmente como se hace con una imagen visual. Un concepto se exploraría, al igual que en el caso de una imagen visual, mediante una secuencia de cambios de atención consecutivos en los que en cada momento se presta atención a alguna de sus características: pone huevos, vive en grupos (ver Ramos, en preparación para más detalles).

Finalmente, el problema del *binding* tampoco ha sido tratado adecuadamente por los modelos de tipo interactivo y ésta es una de las causas principales (junto con la falta de composicionalidad) de que no puedan tratar más que problemas de juguete o que no hayan sido nunca implementados. En cambio, los modelos que se ajustan a las características que hemos descrito anteriormente no solamente pueden tratar problemas de la complejidad de los problemas reales, sino que lo hacen con un rendimiento excelente (parecido al humano) en algunas de las tareas.

Por último, en Ramos (en preparación) se muestra que es posible elaborar un modelo de *priming* semántico que pueda enmarcarse en esta teoría más general, respetando las restricciones impuestas por ella. El modelo resultante explica de forma más adecuada y completa que los modelos actuales de *priming* los datos empíricos obtenidos hasta ahora en las distintas tareas de *priming* y resuelve las paradojas experimentales y las discusiones hasta ahora planteadas. Las restricciones que impone esta teoría más general también deberían respetarse en el caso del acceso léxico y la desambiguación sintáctica, y en este trabajo en preparación se apunta cómo se podría hacer de una forma unificada.

REFERENCIAS

- Alegre, M.A. (2004). *A neural network approach in tasks of natural language processing*. Ph.D. Thesis. Universitat Politècnica de Catalunya, Barcelona.
- Alegre, M.A., Sopena, J.M., & Lloberas, A. (1999). PP-attachment: A Committee machine approach. *Joint SIGDAT Conference on empirical methods in NLP and Very Large Corpora*, 231-238.
- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: activation limits on retrieval. *Cognitive Psychology*, 30, 221-256.
- Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, 139, 201-208.

- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.
- Berwick, R. & Weinberg, A. (1982). Parsing efficiency, computational complexity, and the evaluation of grammatical theories. *Linguistic Inquiry*, 13, 165-191.
- Bienenstock, E., Geman, S., & Potter, D. (1997). Compositionality, MDL priors, and object recognition. En M. C. Mozer, M. I. Jordan & T. Petsche (Eds.), *Advances in neural information processing systems*, vol. 9 (pp. 838-844). Cambridge, MA: MIT Press.
- Brown, G.D.A., Hulme, C., & Preece, T. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107, 127-181.
- Charniak, E. (1997). *Statistical parsing with a Context-Free Grammar and word statistics*, Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI Press/MIT Press, Menlo Park, Ca.
- Chein, J. M., Ravizza, S. M., & Fiez, J. A. (2003). Using neuroimaging to evaluate models of working memory and their implications for language processing. *Journal of Neurolinguistics*, 16, 315-339.
- Churchland, P. & Sejnowsky, T. (1992). *The Computational Brain*. Cambridge, MA: MIT Press.
- Cohen, A. & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 650-663.
- Collins, M.J. (1996). *A new statistical parser based on bigram lexical dependencies*. In Proceedings of the 34th Annual Meeting of the ACL.
- Collins, M.J., & Brooks, J. (1995). *Prepositional phrase attachment*. Proceedings of the 3rd Workshop on Very large Corpora, 27-38.
- Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.
- Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, 1, 123-132.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Davelaar, E.J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, 112, 3-42.
- Desimone, R., Ungerleider, L. G. (1989). Neural mechanisms of visual processing. En Boller & Grafman (Eds.), *Handbook of Neuropsychology*, vol. 2 (pp. 267-299). Amsterdam: Elsevier.
- Desmedt, J.E., & Tomberg, C. (1994). Transient phase-locking of 40 Hz electrical oscillations in prefrontal and parietal human cortex reflects the process of conscious somatic perception. *Neuroscience Letters*, 168, 126-129.
- D'Esposito, M. D., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: Evidence from event-related fMRI studies. *Experimental Brain Research*, 133, 3-11.
- Edelman, S., & Intrator, N. (2003). Towards structural and systematicity in distributed, statically bound visual representations. *Cognitive Science*, 27, 73-109.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1, 1-47.
- Fillmore, C. J. (1988). The Mechanisms of Construction Grammar. *Berkeley Linguistics Society*, 14, 35-55.
- Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Foos, P.W., & Wright, L. (1992). Adult age differences in the storage of information in working memory. *Experimental Aging Research*, 18, 51-57.
- Frege, G. (1903). *Grundgesetze der Arithmetik, Begriffsschriftlich abgeleitet*, 2. Band. Jena.
- Fuster, J.M. (1995). *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*. Cambridge, MA: MIT Press.
- Garavan, H. (1998). Serial attention within working memory. *Memory and Cognition*, 26, 263-276.
- Goldberg, A. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldman-Rakic, P.S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational memory. In V.B. Mountcastle, F. Plum, & S.R. Geiger(Eds.), *Handbook of neurobiology* (pp. 373-417). Bethesda, MD: American Physiological Society.
- Goodale, M.A. & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 20-25.
- Gray, C.M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. USA*, 86, 1698-1702.
- Gupta P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, 33, 1001-1016.
- Harrison, M. (2005). *Discovering compositional structures*. Ph.D. Thesis, Brown University, Providence, RI.
- Henson, R. (2001). Neural working memory. En J. Andrade (Ed.), *Working memory in perspective* (pp. 151-173). Philadelphia, PA: Psychology Press.

- Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics*, 19 (1), 103-120.
- Hummel, J.E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480-517.
- Hummel, J.E., & Holyoak, K.J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220-264.
- Hummel, J.E., & Stankiewicz, B. J. (1998). Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition*, 5, 49-79.
- Joshi A. (1999). *Explorations of a domain of locality*. CLIN'99. Utrecht.
- Kane, M.J., & Engle, R.W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47-70.
- Klapp, S.T., Marshburn, E.A., & Lester, P.T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, 112, 240-264.
- König, P., & Engel, A.K. (1995). Correlated firing in sensory-motor systems. *Current Opinion in Neurobiology*, 5, 511-519.
- Lisman, J.E., & Idiart, M.A.P. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science*, 267, 1512-1515.
- Lloberas, A. (en preparación). *Language processing: A mapping approach*. Ph. D. Thesis, Universitat de Barcelona.
- López-Moliner, J. (1998). *Un enfoque neuronal para la desambiguación del significado*. Unpublished doctoral dissertation. Universitat Politècnica de Catalunya, Barcelona.
- Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279-281.
- McDonald, M.C., Pearlmuter, N.J. & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
- Miikkulainen, R. (1996). Subsymbolic case-role analysis of sentences with embedded clauses. *Cognitive Science*, 20, 47-73.
- Miyake, A., & Shah, P. (1999). (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, England: Cambridge University Press.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 411-421.
- Oberauer, K., Demmrich, A., Mayr, U., & Kliegl, R. (2001). Dissociating retention and access in working memory: An age-comparative study of mental arithmetic. *Memory and Cognition*, 29, 18-33.
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6, 97-107.
- Ramos, P.J. (en preparación). *Memoria de trabajo y procesamiento del lenguaje*. Tesis doctoral, Universitat de Barcelona.
- Ruchkin, D.S., Grafman, J., Cameron, K., & Berndt, R.S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences*, 26, 709-777.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing. volume 2: Psychological and biological models*. Cambridge, MA: MIT Press.
- Shastri, L. & Ajanagadde (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic binding. *Behavioral and Brain Sciences*, 16, 417-494.
- Smith, E.E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283, 1657-1661.
- Sopena, J.M., & Alegre, M. (2000). *ANNP: a neural network parser for real world texts*. International Conference on Pattern Recognition, Catalonia, Spain.
- Sopena, J.M. & Lloberas (en preparación). *Language processing: A mapping approach*.
- Sopena, J.M., Díaz, C., Lloberas, A., & Moliner, J. (1998). *A connectionist approach to prepositional phrase attachment*. In COLING-ACL 98, 1233-1237.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.
- Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovín, & Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* 373, 515-518.
- van Essen, D., Anderson, C.H., & Felleman, D.J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science*, 24, 419-423.
- van Valin, R. D., & Randy LaPolla, (1997) *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.
- von der Malsburg, C. (1981). The correlation theory of brain function. (*Internal Report 81-2*). Göttingen, Germany: Max-Planck-Institute for Biophysical Chemistry, Department of Neurobiology. Reimpresso (1994) en E. Domany, J.L. van Hemmen, & K. Schulten, (Eds.). *Models of Neural Networks II*. Berlin: Springer.
- Zeki, S. (1978). Functional specialization in the visual cortex of the monkey. *Nature*, 274, 423-428.