



UNIVERSITAT DE  
BARCELONA

## Continuidad y cambio: modelización de los aspectos dinámicos de la cognición

Manel Viader Junyent




Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – SenseObraDerivada 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – SinObraDerivada 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-NoDerivs 4.0. Spain License.**

CONTINUIDAD Y CAMBIO: MODELIZACIÓN DE LOS ASPECTOS  
DINÁMICOS DE LA COGNICIÓN

---

  
M. VIADER

Tesis Doctoral presentada por  
Manel Viader Junyent

Director: Dr. Jaume Arnau Gras

BIBLIOTECA DE LA UNIVERSITAT DE BARCELONA



0700296862

Desitjo agrair el seu ajut en la realització d'aquesta tesi a les següents persones:

Al Dr. Jaume Arnau, per haver acceptat la responsabilitat de direcció de la Tesi i pel seu recolzament en la realització del treball.

Al Dr. Josep Ma. Domènech i a tots els companys de la sub-àrea d'Estadística del Departament de Metodologia de les Ciències del Comportament, per rellevar-me d'algunes de les meves obligacions amb l'assignatura. El temps que això m'ha permès guanyar ha estat fonamental per a la realització d'aquesta Tesi.

A la Professora Ma. Luisa Honrubia, pels seus comentaris i orientacions respecte al contingut d'aquest treball.

Al Dr. Joan Ma. Malapeira, per les seves apreciacions, sempre profundes, sobre el meu treball i per la riquesa d'idees que sempre l'envolta.

Finalment, i de manera molt especial, a Pilar Fernández, per estar amb mí en circumstàncies especialment difícils.

## INTRODUCCION

Comenzaré esta introducción pidiendo disculpas por tomarme dos libertades: En primer lugar, y a diferencia del resto de la Tesis, utilizaré en este apartado la primera persona del singular, puesto que considero que se trata de la fórmula más adecuada para expresar algunas ideas en términos menos "neutros" o asépticos. Esto no significa, naturalmente, que en el resto de este trabajo no comprometa numerosas opiniones personales. Pienso que, de no ser así, no podría hablarse de Tesis Doctoral sino, más bien, de una recopilación o revisión más o menos exhaustiva de trabajos anteriores. La segunda licencia se refiere al hecho de introducir algunos breves apuntes autobiográficos que, a mi entender, aclaran algunas cuestiones relacionadas con la presente Tesis.

En concreto, quisiera hacer referencia a mi tesina de Licenciatura, que versaba sobre la modelización matemática de los procesos de adquisición de conceptos. Este trabajo se basaba esencialmente en la utilización de estructuras de transición markovianas para modelizar el proceso de prueba de hipótesis que supuestamente subyace al tipo de tareas considerado. A pesar de que el resultado del trabajo fue satisfactorio (al menos, en opinión de quienes

tuvieron la función de juzgarla), me planteé posteriormente algunas dudas. Aquellos modelos parecían ser "demasiado regulares", a pesar de que estaban ideados para aprehender un fenómeno esencialmente discontinuo (el aprendizaje todo o



nada de un concepto por prueba de la hipótesis correcta). Sin embargo, el sistema cognitivo parece presentar otras discontinuidades de tipo más global; puede pensarse, por ejemplo, en los clásicos fenómenos estudiados por los psicólogos de la Gestalt. Por otra parte, ciertas lecturas sobre las limitaciones de la "causalidad markoviana" (esencialmente algunos artículos de Patrick Suppes) contribuyeron a modificar en algunos aspectos mis puntos de vista sobre el funcionamiento del sistema cognitivo. Por otra parte, algunos compañeros de Facultad que habían trabajado en el ámbito de los modelos markovianos discretos parecían estar más interesados en las versiones continuas del planteamiento de Markov.

Posteriormente, y junto a los profesores Jaume Arnau, Joan Ma Malapeira y Ma Luisa Honrubia, nos planteamos la preparación de una serie de Comunicaciones respecto a las formas de modelización en el ámbito de la psicología cognitiva. Una de las conclusiones inequívocas de aquellos trabajos fue la de que la actividad teorizadora de la mayoría de los psicólogos cognitivos se reducía a la formulación de micro-modelos de ámbito limitado a tareas experimentales muy específicas. Aquella proliferación un tanto incontrolada de modelos planteaba serias dudas sobre la pertinencia de ese tipo de trabajo de investigación o, al menos, sugería la necesidad de analizar en profundidad los presupuestos de los cuales se partía. Estos presupuestos son, en gran parte, los propios

del enfoque computacional-discreto que analizo en la presente Tesis.

Inicialmente, esta Tesis debía estar dedicada al análisis del proceso de solución de problemas en sujetos humanos. Un análisis, siquiera superficial, de este campo de trabajo pone de manifiesto casi inmediatamente la existencia de dos orientaciones contrapuestas: Por una parte, los trabajos clásicos de la teoría de la Gestalt (y de autores relativamente aislados como Maier), que ponían el énfasis en un conjunto de procesos de carácter dinámico tales como campos de fuerzas, tensión del sistema, re-estructuración, etc.; por otra parte, todo el enfoque computacional que parte de los trabajos del grupo de Carnegie-Mellon. La constatación de este hecho me confirmó en la necesidad de valorar más profundamente los fundamentos de las dos orientaciones, y me sugería el posible interés de recuperar algunas ideas relevantes de los teóricos de la Gestalt.

Como sucede en muchas ocasiones, el planteamiento de estas cuestiones y la adquisición de información sobre ellas provocó nuevas dudas. Parecía cada vez más claro que el enfoque computacional-discreto presentaba problemas globales que exigían respuestas también relativamente generales. Algunas de estas respuestas empezaron a esbozarse en el momento en que tomé contacto, casi simultáneamente, con dos líneas de investigación que considero fundamentales: en el campo de la psicología, la formulación de modelos distribucionales del sistema cognitivo; en el ámbito de la matemática (y de la

metamatemática?) el desarrollo de ciertas aproximaciones a la modelización de la discontinuidad que han culminado, al menos provisionalmente, en la eclosión de la teoría de las catástrofes a partir de los años 70 (aunque sus antecedentes se remontan a principios de siglo). La toma de contacto con este tipo de análisis determinó el cambio definitivo de orientación de la Tesis. Esta no se destinaría ya al estudio de la solución de problemas, sino al planteamiento de cuestiones de tipo más general.

A partir de aquí, los acontecimientos se han desarrollado casi por sí mismos. Pienso que la mejor forma de explicar su evolución es referirse directamente a la concatenación lógica de los distintos aspectos de la Tesis. En términos generales, su contenido es el siguiente:

El capítulo 1 pretende realizar una revisión relativamente aséptica y descriptiva de los aspectos principales de lo que, a lo largo del trabajo, denomino "enfoque computacional-discreto". Se trata, esencialmente, de extraer de la amplia bibliografía existente sobre la temática los aspectos fundamentales referidos tanto a las presuposiciones centrales de este enfoque como a los antecedentes que le pueden ser atribuidos. El objetivo es, en todo caso, mostrar cómo las suposiciones básicas que adoptan los teóricos computacionales se derivan lógicamente de tales antecedentes, y realizar al propio tiempo una primera aproximación a algunas de las consecuencias que se deducen de su postura.

El capítulo 2 está dirigido al análisis crítico de los aspectos básicos introducidos en el capítulo anterior. Pienso que el enfoque computacional discreto se enfrenta a dificultades muy considerables, algunas de las cuales son analizadas con cierto detalle. Estas dificultades se refieren, por ejemplo, al lugar que ocupa este tipo de modelos en el marco general de una teoría del procesamiento de información (en concreto, intentaré mostrar que se trata de modelos de la competencia), a las limitaciones que implica la utilización de un código representacional de base proposicional y a las dificultades generales que este tipo de formalismos encuentran para aprehender aspectos tan fundamentales del comportamiento como son su carácter adaptativo y cambiante. En un plano algo más particular, se analizan también los problemas a los que se enfrentan las redes semánticas asociativas en su intento de aproximación al significado. He dejado de lado deliberadamente algunas formulaciones generales como son los conocidos "teoremas limitativos" del tipo teorema de Gödel (aunque sí hago alguna breve mención a la cuestión). El problema de este tipo de teoremas es la dificultad en su interpretación; por ejemplo, el teorema de Gödel ha sido propuesto tanto como demostración de la imposibilidad de construir "máquinas pensantes" (por ejemplo, en el trabajo clásico de Lucas) como para apoyar posiciones mecanicistas (en estudios recientes, por ejemplo, de J. Webb y otros filósofos).

El capítulo 3 ha llevado asociados dos problemas básicos. El primero es el de su extensión. Mi idea inicial era realizar

una revisión lo más exhaustiva posible de los aspectos de otras ciencias que pudieran ser relevantes para el análisis que intentaba realizar. El capítulo partía de la base de que estamos en condiciones de abandonar las consideraciones sobre el carácter "importador" de ideas teóricas que se ha atribuído a la psicología, y se proponía demostrar cómo la evolución de la ciencia contemporánea conduce a una cierta aproximación no reduccionista entre "ciencias naturales" y "ciencias humanas". Sin embargo, el propio enunciado de este objetivo revela ya su enorme complicación. A partir del material que he podido obtener, creo poder afirmar que la realización de la tarea planteada supera amplísimamente los límites de este trabajo. En consecuencia, opté por una vía opuesta a la que planteaba al principio: reducir y condensar al máximo el capítulo con el fin de centrarse únicamente en los aspectos fundamentales.

El segundo problema radicaba en su ubicación. En realidad, el capítulo podría ser emplazado casi en cualquier posición. La opción adoptada responde a una doble consideración: En los dos primeros capítulos planteo una problemática surgida indudablemente en el ámbito de la psicología; queda claro, entonces, que no se trata de cuestiones "importadas" de otras disciplinas. Los capítulos 4 y 5 intentan elaborar una aproximación a los procesos cognitivos alternativa al análisis en términos computacionales, y en este empeño me parecen muy relevantes algunas analogías con la problemática planteada en ciencias como la física (o la "nueva física") y la



biología. En este sentido, parecía totalmente lógico que estas analogías se basaran, al menos en parte, en una exposición previa de algunos aspectos de aquellas disciplinas.

El capítulo 4 se define en gran parte por referencia a su título. Se trata, efectivamente, de sentar las bases para la formulación de la "máquina de Boltzmann" como modelo para el sistema cognitivo humano. Las propiedades fundamentales de la "máquina de Boltzmann" consistirían en el carácter distribucional e interactivo del funcionamiento de las redes de unidades elementales. Esta arquitectura funcional, relativamente simple, posee un sorprendente poder emergente, que la capacita para enfrentarse con éxito a la aprehensión de procesos cognitivos complejos, entre ellos los procesos de abstracción. Finalmente, analizo la presencia de propiedades globales de este tipo de redes, especialmente en lo que se refiere a la propiedad de resonancia. Esta tipo de propiedades resulta fundamental, puesto que puede demostrarse la posibilidad de analizarlas en términos topológicos, lo cual justificará posteriormente el recurso a la teoría matemática de la estabilidad estructural.

Finalmente, el capítulo 5 intenta delinear las condiciones básicas para la aplicación de la teoría de las catástrofes al ámbito de la psicología. El objetivo del capítulo no es propiamente la formulación de modelos específicos (aunque en el apartado 5.3 se hacen algunas sugerencias), sino valorar las posibilidades generales de aplicación de la teoría. Esta precisión me parece importante, puesto que una de las

críticas fundamentales que se dirigen a menudo contra la teoría de las catástrofes se refiere especialmente a su aplicación no siempre rigurosa. La utilización de la teoría como instrumento de modelización será la consecuencia lógica de la valoración de un conjunto de condiciones como, por ejemplo, la posibilidad de considerar el sistema cognitivo humano como análogo a un sistema gradiente.

## INDICE

1. Elementos básicos del enfoque computacional-disco 11
  - 1.1 Plano cognitivo y computación
  - 1.2 Algunos antecedentes: racionalismo y formalismo
  - 1.3 ¿Un lenguaje de la mente?
  - 1.4 Modelos basados en sistemas de producción
  
2. Símbolos y procesos: Bases para un análisis crítico de los modelos computacionales discretos 61
  - 2.1 Niveles de una teoría general del procesamiento de información
  - 2.2 Estructura y naturaleza de los símbolos
  - 2.3 El "cambio cognitivo" desde una perspectiva computacional
  - 2.4 Redes semánticas asociativas
  
3. El cambio desde la perspectiva general de la ciencia 135
  
4. De la máquina de Turing a la máquina de Boltzmann 157
  - 4.1 Modelos distribucionales interactivos
  - 4.2 El poder emergente de los modelos distribucionales
  - 4.3 Propiedades globales de las redes de unidades en paralelo
  
5. Hacia una teoría de la estabilidad estructural en psicología 229
  - 5.1 La teoría de las catástrofes como análisis de la estabilidad estructural: Elementos básicos
  - 5.2 ¿Matemáticas para la psicología?
  - 5.3 Hacia la construcción de modelos del cambio en psicología



1. Elementos básicos del enfoque  
computacional - discreto

---

## 1.1 Plano cognitivo y computación

Cualquier intento de construcción de una psicología cognitiva parte, al menos en principio, de la base de que una adecuada explicación de la conducta requiere la utilización de un vocabulario diferente al que es propio de las explicaciones físicas o biológicas. Ya a un nivel puramente informal e intuitivo es fácil encontrar razones que justifiquen la insuficiencia de tales niveles explicativos. Es suficiente, en general, con plantearse cualquier situación cotidiana y analizar el comportamiento exhibido por las personas que intervienen en esa situación: por ejemplo, ¿cómo puede explicarse el hecho de que una persona cambie el canal de su televisor cuando se ofrece algún espacio que no es de su agrado? o, tomando un ejemplo citado con mucha frecuencia, ¿por qué una persona llama por teléfono a los bomberos al contemplar un incendio?. Un análisis estrictamente físico explicaría la trayectoria seguida por el sujeto en dirección al teléfono y los movimientos que se efectúan al realizar la llamada en términos de fuerzas aplicadas y transferencia de energía. Sin embargo, este nivel de análisis parece claramente insatisfactorio. Lo mismo podría decirse de un estudio en términos de tensión o relajación de músculos, e incluso de un análisis neurofisiológico detallado. Ninguno de estos niveles de análisis, argumentan los psicólogos cognitivos, puede explicar por qué el sujeto realiza esa secuencia de acciones.

Como señala Pylyshyn (1980), en el actual estadio de desarrollo de la psicología cognitiva existe cierta tendencia a complementar estos niveles de estudio con un análisis en el cual intervienen ciertos términos intencionales derivados, en última instancia, de los trabajos de Franz Brentano (McCallister, 1976, ofrece un detallado análisis de la influencia de Brentano en la filosofía y psicología actuales). Diríamos entonces que la secuencia de acciones del sujeto debería ser analizada como un conjunto de actos que incluiría la percepción de un hecho (el incendio), el reconocimiento y clasificación de ese hecho como un accidente, la inferencia de que ese accidente puede causar daños, la decisión de pedir ayuda, el análisis del conocimiento de situaciones similares y, en definitiva, la acción de telefonar con la intención de obtener auxilio. Aunque este vocabulario puede ser insatisfactorio en primera instancia, su utilización parece difícil de evitar, al menos como primera aproximación, dada la orientación actual de la mayoría de los autores que trabajan en psicología cognitiva (Pylyshyn, 1986).

En términos más precisos, los teóricos de la psicología cognitiva (o, al menos, los de orientación computacional), sugieren que la utilización de términos como los anteriores y de la noción de representación en psicología permite dar cuenta de ciertas regularidades o generalizaciones de la conducta que no pueden ser aprehendidas por análisis meramente físicos, neurofisiológicos e incluso fenomenológicos

(Fodor, 1968, 1975, 1981; Pylyshyn, 1980; Dennett, 1978, etc.). Tras esta idea se encuentra un concepto de explicación según el cual una aproximación adecuada a la conducta no se agota en la descripción de algunas (e incluso todas) sus causas, sino que exige la aprehensión de "generalizaciones predictivas" que permitan englobar todo lo que de sistemático tiene el comportamiento humano. Se trata por tanto, de aprehender, en la medida de la posible, el grado de sistematicidad y regularidad que presenta el comportamiento de los sujetos humanos; diferentes vocabularios descriptivos, entre ellos el propio de la psicología cognitiva, permiten aproximarse a generalizaciones distintas y, de alguna forma, complementarias.

La cuestión es lo bastante importante como para exigir una mayor concreción. Obviamente, algunos procesos (en cierto sentido, todos) pueden ser explicados perfectamente analizando la conexión causal entre algunas de sus propiedades físicas, de forma que al alterar algún parámetro físico se producen cambios en algún otro parámetro de acuerdo con una ley que conecte las dos propiedades consideradas. Esta causalidad física no puede, sin embargo, explicar las conexiones que son independientes de las propiedades físicas particulares del input. Retomando el ejemplo que se utilizaba algunas líneas más atrás, un sujeto puede darse cuenta de la existencia de un incendio de muy diversas formas: viendo las llamas, observando la presencia de una gran cantidad de humo, oyendo exclamaciones de otras personas, etc. Sin

embargo, este sujeto actuará presumiblemente de forma muy semejante en cada uno de estos casos (por ejemplo, llamará por teléfono para pedir ayuda). Obviamente, un análisis a nivel meramente físico no puede dar cuenta de esta generalización, puesto que las diversas señales físicas del fuego no tienen prácticamente nada en común. La generalización sólo puede ser aprehendida, de acuerdo con los teóricos computacionales, postulando la existencia de estados de "creencia", "suposición", "objetivo", etc., es decir, de lo que ha sido denominado "actitudes proposicionales" (Fodor, 1975, 1981a, 1983). De esta forma, los estados mentales serían distinguibles entre sí tanto en términos del contenido de las representaciones asociadas a los mismos como de la relación establecida con esas representaciones (saber, creer, desear, dudar...).

Pylyshyn (1981) señala que la existencia de relaciones con la representación (es decir, de actitudes proposicionales) es crucial para el marco teórico de la psicología computacional, puesto que introduce la posibilidad de aprehender ciertas propiedades del comportamiento. Supongamos que un sujeto presenta cierta conducta caracterizada por una función  $f_1$  cuando el sujeto cree algo, y por otra función  $f_2$  cuando esa creencia es distinta. Tomemos el ejemplo (uno de los más apreciados por Pylyshyn en su polémica con los teóricos de la imagen) de una persona que está realizando la rotación mental de un objeto. Supongamos también que la función  $f$  que describe el comportamiento del sujeto presenta

cierta relación lógica con el contenido de una creencia del sujeto: por ejemplo, esa persona puede creer que el objeto que está siendo rotado es de gran peso y, en consecuencia, que es muy difícil someterlo a grandes aceleraciones; el resultado de esta creencia es que la función que relaciona la latencia de respuesta con el ángulo de rotación será distinta a la que se hubiera obtenido si el sujeto creyera que el peso del objeto a rotar es pequeño.

En su análisis de este fenómeno de "penetrabilidad cognitiva" Pylyshyn (1973, 1981, 1983) argumenta que los procesos realizados sobre la información se llevan a cabo sobre la base de unidades de interpretación, de un "conocimiento tácito" (Fodor, 1968, 1981a) de carácter conceptual, y sugiere que cualquier explicación basada en una supuesta funcionalidad de las imágenes mentales o en propiedades de un medio análogo no será capaz de dar cuenta del fenómeno a explicar. El problema es exactamente el mismo que se exponía al comentar el ejemplo del incendio: no podemos limitarnos a postular la existencia de un conjunto de leyes físicas o biológicas que produzcan la función observada  $f$ , puesto que la regularidad del comportamiento del sujeto depende del contenido semántico (en este caso, la creencia) de la representación y de las relaciones lógicas establecidas en ella; en consecuencia, cualquier explicación no establecida en este "nivel cognitivo" (y, en particular, las que se remitan a las propiedades de un medio representacional que presente analogías con el medio físico) será insuficiente. Aunque en



cada caso particular la conducta será causada por algunos procesos físicos, biológicos y neurofisiológicos, el principio explicativo general se encuentra en un nivel cognitivo susceptible de aprehender la generalidad de la conducta observada. En la terminología de Fodor (1975), existe una reducción de señales (inscripciones físicas) a principios físicos, pero no se da una reducción semejante para los tipos de vehiculos de representación (conjuntos de señales distintas que se refieren a un mismo simbolo).

La distinción entre señales o inscripciones físicas y tipos de vehiculos de representación es fundamental para el enfoque computacional de la psicología. En efecto, es fácil darse cuenta de que la semántica de las representaciones no puede en si misma causar la conducta; solamente la forma material de la representación puede actuar como causa eficiente. Cuando decimos que el sujeto A realiza la conducta X a causa de que cree P estamos "atribuyendo causalidad" a una representación pero, obviamente, se trata únicamente de una utilización poco precisa del lenguaje que no implica en absoluto la consideración de la representación como causa eficiente. Una afirmación de este tipo equivaldría a una vuelta a la "causalidad mental" propia de las distintas formas de dualismo. Lo que realmente causa la conducta son ciertas propiedades físicas del estado representacional aunque, por supuesto, esas propiedades deben mantener algún tipo de relación con el contenido de la representación. ¿Cómo se establece esa relación?; ¿cómo es posible que las conductas

causadas por ciertas propiedades físicas sean equivalentes a las que pueden ser aprehendidas por referencia al contenido semántico de las representaciones?. En la respuesta a estas cuestiones se encuentra una de las claves del enfoque computacional.

La idea que proponen los teóricos computacionales es bien conocida: lo que hace el cerebro es exactamente lo que hacen los ordenadores cuando computan funciones numéricas (Pylyshyn, 1986). La conducta de los ordenadores (y, supuestamente, también la de los seres humanos) es causada por las propiedades físicas de ciertas clases de sub-estados que corresponden a códigos simbólicos. Estos códigos reflejan todas las distinciones semánticas necesarias para hacer que la conducta corresponda a las regularidades expresadas en los términos semánticos. Utilizando otros términos, los símbolos (tipos) son clases de equivalencia de propiedades físicas capaces al mismo tiempo de causar la conducta y de preservar ciertas interpretaciones y distinciones entre los términos semánticos. En consecuencia, estar en un cierto estado representacional será equivalente, de acuerdo con el enfoque computacional, a poseer cierta expresión simbólica activa en la memoria.

El requisito de preservación de distinciones semánticas es válido para cualquier teoría de la representación (Anderson, 1979): a objetos distintos (o, al menos, a inputs sensoriales distintos) deben corresponderles representaciones distintas. Más aún, como señala el propio Anderson en su



conocido argumento de indecibilidad, es posible construir teorías representacionales diferentes que preserven las mismas regularidades semánticas y, en consecuencia, sean indistinguibles a nivel conductual (desde el momento en que tales regularidades semánticas corresponden, de alguna forma, a sucesos físicos que constituyen la causa eficiente de la conducta).

Fodor (1981a) ha expresado algunas de las ideas anteriores en términos muy precisos: una actitud proposicional respecto a P es una relación computacional con la representación de P. La fuerza de la noción de relación o transformación computacional radica precisamente en el hecho de que se trata de un tipo de transformación que preserva ciertos tipos de relaciones semánticas tales como confirmación, implicación lógica, etc. (Fodor, 1975; Haugeland, 1981). Más aún, el círculo se cierra cuando nos damos cuenta de que tales relaciones sólo se verifican, por definición, entre aquellas categorías de objetos a las que puede adscribirse contenido proposicional, es decir, entre aquellas categorías de objetos de las que puede decirse, entre otras cosas, "significa que P" (Fodor, 1983). Esta idea es fundamental, por ejemplo, para la teoría del lenguaje propuesta por Chomsky (1965, 1980). Siguiendo la tradición cartesiana (que examinaremos más adelante con algún detalle), lo que Chomsky (1980) considera como innato no son únicamente ciertos mecanismos psicológicos, sino también un cierto corpus de información: el niño nace conociendo (o "cognosciendo", usando el neologismo

propuesto en la traducción de la obra de Fodor, 1983) ciertos hechos referidos a universales lingüísticos. Es decir, lo innato posee contenido proposicional, por lo cual en el desarrollo cognitivo lo que viene dado de forma endógena se desplegará de modo computacional (Fodor, 1983).

Fodor (1981a) presenta un análisis sistemático de la distinción que puede establecerse entre las suposiciones comunes a las diversas teorías representacionales de la mente y las suposiciones, considerablemente más fuertes, propias de los modelos computacionales. En particular, los procesos computacionales poseen un carácter simbólico, puesto que no se definen sobre objetos sino sobre representaciones, y formal, desde el momento en que su aplicación a esas representaciones se realiza en virtud de la sintaxis de la mismas y, por tanto, sin referencia a propiedades semánticas de la información tales como verdad, referencia o significado. No hay contradicción con lo expuesto anteriormente, puesto que las distinciones semánticas son tratadas, desde el punto de vista computacional, en términos de condiciones formales de las representaciones, de forma que las diferencias semánticas se reducen a distinciones sintácticas (Fodor, 1975, 1980, 1981a; Pylyshyn, 1980; Stich, 1983). En otros términos, los estados mentales son analizados, desde esta perspectiva, como relaciones entre representaciones, aunque no deba descartarse la posibilidad de analizar la relación de las representaciones con el mundo externo (Fodor, 1981a). Esta última tarea quedaría encomendada a una hipotética

psicología naturalista, sobre cuya viabilidad el propio Fodor manifiesta un considerable escepticismo (Fodor, 1981a, pp. 233 y ss.), aun siendo consciente del peligro de que el solipsismo metodológico, asumido como estrategia para la individualización de los procesos mentales, abra de hecho el camino a un solipsismo epistemológico difícilmente asumible como actitud científica.

El énfasis en los aspectos sintácticos o formales de la representación es el resultado lógico del planteamiento de la causalidad que realiza el enfoque computacional. Las propiedades causales de los estados psicológicos (que radican, obviamente, en el plano físico) determinan sus propiedades formales para los propósitos de la psicología cognitiva; las mismas relaciones formales pueden ser establecidas a partir de estados físicos diferentes. En otras palabras, las relaciones formales son relaciones causales descritas computacionalmente. Si dos tipos de vehículos de representación A y B están computacionalmente conectados, entonces estarán también conectados en su implementación material. Las relaciones computacionales entre estos símbolos-tipo serán entonces, en un importante sentido, puramente sintácticas, y no estarán afectadas por los rasgos semánticos de la representación (Maloney, 1985). Cuando los teóricos computacionales hablan del contenido de la representación se refieren a su contenido conceptual (objetivos, creencias, etc.), es decir, a relaciones que han definido previamente como computacionales. El contenido será función de la forma,

en el sentido de que dos estados mentales tendrán contenidos distintos solamente si presentan relaciones computacionales con cadenas de simbolos formalmente distintas. De esta forma, la referencia puede ser tratada en términos de una función semántica definida recursivamente sobre una estructura de cadenas de simbolos, al estilo de la semántica tarskiana, en la cual el significado de las sentencias es determinado, en el marco de un cálculo formal, en base a las propiedades combinatorias de las cadenas de simbolos que representan esas sentencias (Tarski, 1956; Pylyshyn, 1979, 1986).

La condición de formalidad de los procesos mentales es auténticamente fuerte y de ella se derivan inmediatamente una serie de consecuencias importantes. En la medida en que consideremos los procesos mentales como computaciones debemos aceptar que la repercusión de la información ambiental sobre tales procesos se agota en el carácter formal de lo que los sentidos impriman en la memoria operativa del sistema, sin que tenga importancia el hecho de que esa impresión corresponda o no a un hecho "verdadero". Esto equivale, como señala el propio Fodor (1981a) a considerar los sentidos como una especie de "oráculos", puesto que nos negamos a entrar en la cuestión de si la información que nos proporcionan proviene de un mundo externo objetivo o, por el contrario, es únicamente el resultado de la actividad de los propios sentidos, como postula el idealismo clásico.

De hecho, si se programa a una máquina para que responda de forma adecuada preguntas sobre Historia Universal, esa

máquina no nos proporcionará en absoluto respuestas sobre ese tema. El trabajo de asignar el resultado de la computación a un cierto referente lo realiza el programador, no la máquina. De la misma forma, puede suministrarse a un computador un "entorno" artificial compuesto, por ejemplo, de objetos geométricos (Winograd, 1971). El computador podrá ejecutar un buen número de operaciones aparentemente (?) inteligentes sobre dichos objetos, a pesar de que en esas circunstancias no hay referencia posible al entorno y de que éste en realidad ni siquiera existe, puesto que llega el ordenador via programa y no a través de ningún órgano sensorial.

Por otra parte, las propiedades físicas de los sistemas en los cuales está inserta la representación resultan ser irrelevantes. Para que dos sistemas -por ejemplo un sujeto humano y un computador- crean que "el carbón es negro" no es necesario que sean físicamente semejantes en alguna forma específica, sino que su semejanza debe radicar en un nivel funcional. La reducción de los predicados mentales a predicados físicos es substituida por su reducción a predicados de un sistema de estados lógicos que pueden ser representados, por ejemplo, en una máquina de Turing. El marco conceptual resultante es básicamente una teoría de la identidad en lo referente a las señales físicas -cada suceso mental individual se corresponde con un suceso ocurrido en el soporte físico del sistema- y un funcionalismo en lo referente a los tipos de vehiculos de representación. Este enfoque



funcionalista es utilizado por Fodor (1965, 1981b) para dar respuesta a las insuficiencias que presentan el dualismo y el conductismo en tanto que "filosofías de la mente". En particular, el funcionalismo fodoriano permite eludir todo problema relacionado con la "causalidad mental" (desde el momento en que circunscribe las causas eficientes al mundo físico), y postular al propio tiempo la existencia de estados mentales en un sentido no solamente heurístico, a diferencia del carácter de los enunciados conductuales hipotéticos propios del conductismo lógico (Jacquette, 1985).

## 1.2 Algunos antecedentes: Racionalismo y formalismo

La tendencia formalista exhibida por los teóricos computacionales no es, por supuesto, un fenómeno aislado en el contexto general de la actividad científica y filosófica. Como señala muy acertadamente Rivière (1984, 1986), un antecedente fundamental del enfoque computacional puede encontrarse en la introducción de representaciones analíticas en la geometría a partir de la aritmetización del plano propuesta por Descartes. Más que el ejemplo concreto, lo importante aquí es señalar que la substancial innovación que representa el cambio de la representación analógica presente en la geometría clásica a la representación analítica consustancial al álgebra se enmarca en un programa mucho más amplio cuyo objetivo es la "matematización" del propio pensamiento,

es decir, la consecución de un formato canónico para las operaciones mentales que permitiera, en definitiva, proporcionar un carácter verdaderamente racional a las operaciones del intelecto. El hallazgo de este formato universal es vital para la concepción cartesiana puesto que, de acuerdo con ella, la actividad racional es la fuente básica de aprehensión de la realidad y, por ese mismo motivo, debe estar rigurosamente sujeta a restricciones metodológicas.

En esta línea, Descartes no ambiciona únicamente enriquecer las matemáticas, sino ofrecer vías para la resolución metódica de todos los problemas que pueda plantear la ciencia. Esto exige, en primer lugar, una jerarquización de los distintos órdenes de problemas de acuerdo con su naturaleza. Esta ordenación jerárquica es relativamente simple y natural para los problemas que tratan con hechos discretos (los problemas son, o bien racionales, o bien irracionales, o bien imposibles), pero las dificultades son mayores cuando se trata con realidades continuas. El proyecto de Descartes incluye precisamente la construcción de una jerarquía de problemas en el ámbito de lo continuo, y para ello era necesario romper con la "geometría cinemática" propia, por ejemplo, de Torricelli y Roberval y heredada de algunos matemáticos griegos. En este tipo de geometría, las líneas son generadas por composición de movimientos: por ejemplo, Arquímedes define la espiral como el resultado combinado de la rotación de una semirrecta en torno a su origen y de la traslación de un punto sobre dicha semirrecta a partir de

ese mismo origen. La principal objeción cartesiana a la geometría cinemática se refiere a la inconmensurabilidad de los movimientos que generan las líneas; es decir, los movimientos que engendran una línea son independientes entre sí y, por tanto, es imposible medir la posición de un punto mediante su relación con el desplazamiento de otro punto. El tiempo es el único vínculo entre los dos movimientos. En términos modernos, sería imposible eliminar el parámetro común a los dos desplazamientos para tratar de obtener una ecuación algebraica (de Gant, 1982). Si las matemáticas (y, en última instancia, el propio pensamiento) debían constituirse en una secuencia perfectamente reglamentada de operaciones, el tránsito hacia la geometría algebraica era inevitable. Esta huida de lo continuo será una constante en la línea histórica que conduce desde el cartesianismo hasta el actual enfoque computacional.

En su brillante análisis, Rivière (1986) muestra que el ideal cartesiano, centrado en la consecución de un formato canónico del pensamiento, es también el hilo conductor que da sentido al trabajo de un buen número de filósofos y científicos. Por citar solamente algunos ejemplos ya clásicos en este tipo de análisis, debe mencionarse a Leibnitz y su "arte combinatoria" (un claro precedente de la tendencia "sintáctica" propia del enfoque computacional-discreto); a Boole, con su interpretación lógica del álgebra, fundamental para la matemática posterior; a Frege y sus proposiciones despsicologizadas, que conducirán a la distinción entre



leyes del Pensamiento y leyes del pensar; los intentos del propio Frege (1879) o de Russell (1903) de realizar una reducción lógica de la propia matemática; e, incluso, el propio programa del positivismo lógico cuyo objetivo, de acuerdo con Carnap (que fue discípulo del propio Frege) es elaborar un sistema formal y prescriptivo que permitiese una caracterización formal de todas las ciencias.

Resulta claro que el moderno enfoque computacional en psicología es heredero de la tradición formalista-analítica-discreta que estamos describiendo sucintamente. Como señala muy acertadamente de Vega (1981), al antipsicologismo de los lógicos ha seguido el logicismo de los psicólogos, que han utilizado la lógica no solamente como instrumento formal, sino también, en ocasiones, como heurístico en la formulación de hipótesis psicológicas, desembocando así en un isomorfismo entre procesos mentales y operaciones lógicas que, como veremos en el siguiente capítulo, resulta bastante problemático. Más aún, el dispositivo lógico conocido como máquina de Turing, punto de referencia indispensable para el computacionalismo, debe su existencia al planteamiento de un problema surgido en el contexto de los intentos de formalizar las matemáticas, en oposición a un enfoque más intuicionista. En concreto, en el año 1900 el matemático alemán David Hilbert presentó en el Congreso Internacional de Matemáticos, que se celebraba en París, su conocida lista de problemas no resueltos en el ámbito de la teoría matemática. De hecho, la reflexión de Hilbert le

llevó a plantear un problema general, el Entscheidungsproblem o problema de la decisión. En la más pura línea cartesiana, se trata de descubrir un método general que nos permita determinar si una fórmula de la lógica formal puede o no satisfacerse o declararse verdadera.

El matemático inglés Alan Turing, de la Universidad de Cambridge, abordó el problema planteado por Hilbert en sus justos términos (al menos, desde el enfoque que estamos considerando). Ante todo, debemos plantearnos cómo dar al concepto de método una definición precisa. Si aceptamos la idea intuitiva de que un método es un conjunto de pasos que pueden ser ejecutados mecánicamente, sin intervención creativa alguna, podemos construir un modelo detallado del proceso de análisis de un problema científico en el cual dicho proceso sea descrito mediante una serie de algoritmos que, a su vez, no son sino conjuntos o secuencias de pasos simples. Estos pasos pueden ser ejecutados por dispositivos mecánicos relativamente sencillos que constituyen la concreción material del método de razonamiento matemático propuesto. Sirviéndose de uno de estos dispositivos, Turing concluyó que el segundo problema de Hilbert (demostrar que los axiomas de la aritmética ordinaria son coherentes entre sí) es irresoluble, es decir, no es computable.

En términos generales, si a una máquina de Turing se le da tiempo suficiente puede llevar a cabo cualquier cómputo que pueda realizar un ordenador digital, por potente que éste sea. Por este motivo, la máquina de Turing es un instrumento

conceptual indispensable para el estudio teórico de la capacidad de resolución de problemas que poseen los ordenadores. De hecho, el diseño de los modernos ordenadores digitales no hace otra cosa que sacrificar la claridad de funcionamiento en aras de la rapidez, pero a nivel de estados lógicos el funcionamiento del ordenador y el de una máquina de Turing son perfectamente idénticos. El ordenador posee un número finito de estados internos, y su respuesta a cualquier input viene determinada completamente por el conjunto finito de enunciados que conforman su programación. En consecuencia, el funcionamiento de un ordenador digital puede ser descrito por un número finito (aunque potencialmente muy grande) de dígitos en un sistema binario como el que utiliza la máquina de Turing y, por tanto, esta máquina puede simular el funcionamiento del ordenador. De la misma forma, cualquier ordenador es capaz, si se le proporciona una capacidad de memoria suficiente (capacidad que puede concretarse sencillamente en una cinta de papel movible y legible de longitud suficiente), de simular el comportamiento de cualquier otro ordenador; para ello, el simulador debe ser programado para funcionar como una máquina universal de Turing (Hopcroft, 1984).

La consecuencia fundamental que se deriva de lo que acabamos de exponer es la constatación de que todos los ordenadores digitales pueden computar exactamente la misma clase de funciones matemáticas, a saber, las que sean computables por alguna máquina de Turing. El hecho de que esta clase sea

única refuerza la definición formal de computabilidad formulada por Turing.

Pylyshyn (1986) analiza las repercusiones del trabajo de Turing en el ámbito de la psicología y, en particular, resalta el hecho de que, para derivar su definición de computabilidad, Turing debió introducir las nociones de demostración y deducción en su sistema formal en términos de señales simbólicas (concretamente, marcas en una tira de papel) cuya manipulación, puramente mecánica, es totalmente independiente de la forma en que dichos símbolos sean interpretados. De hecho, la única propiedad que poseen esas señales en este tipo de sistema es de tipo puramente nominal: la identidad de tipo. Si se toman las señales simbólicas como átomos elementales e inestructurados no podrá decirse de ellos que posean un cierto grado de semejanza, o que se parezcan en determinados aspectos y no en otros; simplemente se podrán construir copias indistinguibles de cada señal, que actuarán como marcas del mismo símbolo. A partir del conjunto de átomos elementales podrán postularse un conjunto de reglas sintácticas de relación que hagan posible la construcción de cadenas de símbolos adecuadas para la representación de estados mentales complejos.

La noción de universalidad del mecanismo formal, introducida por los matemáticos en los años 30, constituye una referencia fundamental para el moderno enfoque computacional, puesto que introduce la idea de que un mecanismo formal de manipulación de símbolos es capaz de producir cualquier función

input-output incluyendo, por supuesto, las que son propias de la conducta humana. En su análisis de los sistemas de procesamiento de símbolos, Newell (1980) señala que la afirmación de su presunto carácter universal exige la partición de los inputs del sistema en diversos componentes, uno de los cuales designaría o identificaría una cierta función, que sería ejecutada sobre el componente restante del input, el cual sería tomado a su vez como dato inicial para dicha función. Sin embargo, la noción de universalidad encuentra aquí una dificultad que ya fue prevista en su momento por Turing. Una vez especificadas las propiedades de su máquina en tanto que expresión de un proceso formal de razonamiento, Turing estableció la conexión entre su propio trabajo y los resultados obtenidos por el matemático Georg Cantor. La idea básica de Cantor es que, aunque no exista ningún número entero que sea el máximo, cualquier conjunto infinito de objetos que sean numerables (es decir, que puedan ser puestos en correspondencia biunívoca con el conjunto de los números enteros) es necesariamente un conjunto del mismo tamaño que el de los enteros. Dado que cualquier máquina de Turing que compute una cierta función puede ser expresada como una serie finita de caracteres, todas las posibles máquinas de Turing y, por tanto, todas las funciones computables, pueden ser numeradas y, en consecuencia, puestas en correspondencia con el conjunto de los números enteros. Sin embargo, Cantor demuestra también que existen infinitos conjuntos que no son numerables, en el sentido de que es imposible establecer tal



correspondencia. Entre los ejemplos de conjuntos no numerables se encuentra el conjunto de todas las funciones dependientes de números enteros, con valores también enteros. El número de tales funciones es mayor que el de números enteros (siempre en el sentido de imposibilidad de establecimiento de correspondencia biunívoca con el conjunto de los enteros), de forma que es imposible que existan suficientes programas como para computar toda posible función. Un resultado semejante fue obtenido por Church (1940) a partir de los estudios de lógica de la demostrabilidad iniciados por Kurt Gödel: a partir de un lenguaje formal coherente (el lambda-cálculo, origen del moderno lenguaje LISP, fundamental en inteligencia artificial), Church demuestra la existencia de funciones matemáticas no computables en el sistema (Davis, 1967).

Los estudios de Turing, Godel y Church introducen un límite a la variabilidad de las conductas descriptibles por un sistema formal de manipulación de símbolos. Esto, sin embargo, no representa un problema grave para el enfoque computacional, a no ser que pueda demostrarse la existencia de funciones no computables ejecutadas por sujetos humanos (volveremos brevemente sobre este punto más adelante). Lo que interesa resaltar aquí es que las funciones que interesan al enfoque computacional, si pretende erigirse en teoría explicativa de la conducta, son aquellas que pueden ser ejecutadas de forma exclusivamente mecánica, sin intervención de ningún tipo de inteligencia, puesto que sólo de esta forma

puede ser evitada una regresión homuncular infinita. El enfoque formalista proporciona una caracterización útil de tales funciones: como señala Pylyshyn (1986), durante los últimos 50 años ha sido desarrollado un amplio conjunto de sistemas formales, siempre con la idea, explícita o implícita, de no intervención de un agente inteligente en su ejecución. Todos estos sistemas han mostrado el mismo conjunto de funciones de expresión finita: las "recursivas", o "funciones efectivamente computables". Aunque se han presentado diversas definiciones de función computable, todas ellas van en la línea de afirmar que cualquier comportamiento "descrito de forma precisa" puede ser ejecutado por un programa de ordenador. McCulloch (1965) o von Neumann (1966) definen esta descripción precisa en términos de "descripción verbal rigurosa" o "expresión no ambigua y rigurosa en un número finito de palabras". De hecho, la versión original de esta idea es debida a Charles Babbage, quien afirmaba que su "máquina analítica", antecedente de la máquina de Turing, era "capaz de reproducir todas las operaciones que el intelecto ejecuta a fin de obtener un determinado resultado, siempre que estas operaciones sean a su vez susceptibles de una definición precisa" (Morrison y Morrison, 1961). La idea de recursividad como condición de computabilidad es conocida como "tesis de Church", puesto que fue introducida por este autor en el marco de sus trabajos sobre lógica de la demostrabilidad. La "tesis de Church" es también aceptada explícitamente por algunos teóricos de la imagen (Kosslyn y

Pomerantz, 1977), lo cual, como se verá en su momento, tiene implicaciones importantes sobre el contenido de los modelos elaborados por estos autores.

### 1.3 ¿Un lenguaje de la mente?

Recientemente, un conjunto de importantes trabajos ha analizado exhaustivamente la relación entre el enfoque computacional de la psicología cognitiva y las representaciones de formato proposicional. En nuestro contexto, deben ser citados inexcusablemente los trabajos de Rivière (1984, 1986), en los cuales queda perfectamente claro que la condición fáctica para una definición computacional precisa de los procesos mentales es la existencia de representaciones de tipo discreto y analítico. A este respecto, es necesario resaltar algunas ideas que se desprenden (a veces, de forma un tanto mediata) de la lectura y análisis de algunos textos de los principales teóricos de la psicología computacional. En primer lugar, es importante resaltar que la noción de símbolo atómico discreto es la base de cualquier conocimiento de tipo formal (Pylyshyn, 1973). Para ser más precisos, los símbolos atómicos son imprescindibles para cualquier expresión o cálculo susceptibles de ser expresados mediante alguna notación (esta idea es importante y será retomada en el siguiente capítulo). Como se señaló algunas páginas más atrás, la existencia de símbolos elementales de tipo



discreto, no estructurados y copiables permite la construcción, mediante la aplicación de las reglas sintácticas correspondientes, de cadenas de símbolos capaces de representar estados mentales y, sobre todo, susceptibles de ser tratadas de forma computacional (es decir, de ser procesadas a través de una secuencia de pasos discretos representada por un algoritmo). La noción de comparabilidad y copiabilidad es fundamental, puesto que a partir de ella es posible definir el reconocimiento de muestras del mismo símbolo como operación elemental realizada por la arquitectura funcional del sistema y, en consecuencia, introducir la idea de símbolo-tipo, esencial para todo el marco conceptual que estamos analizando.

¿Cuál es, no obstante, la naturaleza de esas cadenas de símbolos que constituyen la materia prima de las representaciones?. ¿Cuál es el formato que toman las propias representaciones?. Pylyshyn (1973) afirma que es necesario suponer que las representaciones mentales toman un formato proposicional para explicar cómo las personas pueden, por ejemplo, describir dibujos con palabras o crear dibujos para ilustrar un determinado material verbal. Las representaciones deberían poseer, de acuerdo con esto, un carácter "neutral" o amodal que las hiciera útiles como "lenguaje intermedio" entre los distintos formatos que pueden presentar los datos de entrada. Chase y Clark (1972), Anderson y Bower (1973), Fodor (1975, 1978, 1981a), Anderson (1978, 1980) y otros autores han insistido en esta misma idea. Por ejemplo,

Fodor (1978) argumenta sobre la imposibilidad de aprendizaje de un lenguaje cuyos términos expresen propiedades semánticas no expresadas por los términos de algún otro lenguaje previamente conocido. Si aplicamos este argumento al caso del lenguaje natural, resulta evidente que, desde el punto de vista de Fodor, la posibilidad de acceder a su aprendizaje se basa en la existencia de otro lenguaje previo que permita, mediante la aplicación de ciertas reglas de verdad, determinar la extensión de los predicados del lenguaje natural. Si se desea evitar una regresión infinita debe recurrirse a la idea de que este lenguaje previo (el "lenguaje de la mente") posee un carácter innato, puesto que su aprendizaje exigiría a su vez la existencia de otro lenguaje anterior que, por su parte, debería ser innato o aprendido, y así sucesivamente. En consecuencia, debe distinguirse muy claramente entre lo verbal, es decir, el lenguaje natural utilizado con funciones básicamente comunicativas, y lo proposicional, un sistema simbólico interno no accesible a la conciencia y de carácter abstracto (Pylyshyn, 1977; Anderson, 1978).

Esta última afirmación viene avalada, al menos en principio, por un conjunto de fenómenos relacionados con el significado. Puede citarse, por ejemplo, el conocido fenómeno por el cual los sujetos tienden a recordar el significado de las frases, más que su expresión literal (Sachs, 1967; Anderson y Bower, 1973; Kintsch, 1974), fenómenos de integración semántica (Rosenberg y Simon, 1977), efecto de Jarvella,

abstracción de la invarianza de significado entre paráfrasis (Norman y Rumelhart, 1975), etc.

Se admite clásicamente que existen tres características básicas definitorias de una proposición (Frege, 1918-1919/1974): su carácter abstracto, la posesión de un valor de verdad y la existencia de reglas de formación. En cuanto a la primera cuestión, se ha hecho ya referencia a todo un conjunto de fenómenos que parecen apoyar la idea del carácter amodal o supramodal de las proposiciones, referido a su independencia respecto al canal sensorial de entrada de la información. El problema se plantea en cuanto a la conexión o vinculación entre el nivel proposicional abstracto y el funcionamiento efectivo de esos canales sensoriales, puesto que, como señalan de Vega (1981) y Rivière (1986), es necesario, en algún momento, salir del círculo cerrado que representa la definición del significado de un concepto por referencia a otros conceptos y remitirse a elementos de carácter sensorial. Los propios psicólogos de orientación computacional son conscientes de la necesidad de definir la forma que toma la relación entre datos sensoriales y plano cognitivo o, en términos de Pylyshyn (1986), de "construir el puente entre lo físico y lo simbólico". Para ello el propio Pylyshyn define los criterios que, a su juicio, deben definir el funcionamiento de un "transductor psicológico" capaz de realizar la función que estamos considerando. Estos criterios de funcionamiento deberían ser básicamente tres:

a) La función ejecutada por el transductor es primitiva y posee un carácter no simbólico, es decir, forma parte de la arquitectura funcional del sistema, definida como un conjunto de capacidades elementales no simbólicas que dicho sistema posee. El carácter no simbólico del funcionamiento del transductor nos conduce, en principio, fuera del marco de la psicología computacional e intenta romper el círculo vicioso propio de un análisis en términos de relaciones formales entre representaciones. Pero, si se acepta un punto de vista computacional-discreto, se plantea un nuevo problema, puesto que los referentes últimos de la relación del organismo con su entorno pasan a ser entonces elementos sensoriales o pautas motrices que, si no son expresados en un formato proposicional, exigen su propio código o sistema representacional específico (de Vega, 1981). Retomaremos esta importante cuestión en el siguiente capítulo.

b) Un transductor es un componente dirigido por los estímulos y que opera independientemente del sistema cognitivo.

c) El funcionamiento de un transductor puede ser definido como una función entre sucesos físicos y símbolos de forma que: 1. El origen de la función debe ser descrito en el lenguaje de la física; 2. La imagen de la función está compuesta por símbolos atómicos discretos; 3. La transformación desde el input, descrito físicamente, a un suceso computacional, también descriptible en forma física, debe seguir las leyes físicas.

Hablando en términos generales, un transductor produciría un código C (por ejemplo, un n-tuplo de símbolos) siempre que sea estimulado por cierta clase de sucesos del entorno que admitan una cierta descripción física D (sin excluir la posibilidad de que la relación entre D y C esté modulada por el estado del organismo, lo cual introduce un nuevo nivel de complejidad). Puede encontrarse un antecedente relevante (entre otros) de esta idea en el concepto de conversor analógico-digital propuesto por Dretske (1981).

Aunque, evidentemente, las tres características citadas no constituyen sino una declaración de intenciones o de deseos, no nos parece pertinente (al menos, en este punto del trabajo), introducirnos en cuestiones técnicas referidas a la viabilidad y propiedades de un transductor de este tipo. En cualquier caso, lo importante aquí es señalar la significativa preocupación de los teóricos computacionales más lúcidos por encontrar respuestas al peligro de solipsismo epistemológico que subyace a su marco teórico. Aunque en el campo de la lógica, de la cual proviene el concepto de proposición transplantado a la psicología, las proposiciones llevan asociado un valor de verdad, esto no es necesariamente cierto en lo que se refiere a la proposición como entidad psicológica (Anderson, 1978). Rivière (1986) ha analizado profundamente este punto: la concepción lógica de la proposición debe entenderse como una función que proporciona valores de verdad para los estados del mundo; el concepto psicológico de proposición debe ser mucho más flexible, puesto que debe



enfrentarse a enunciados o significados para los cuales no es factible encontrar valores de verdad o que no son asimilables a los modelos lógicos clásicos (de ahí el énfasis de la inteligencia artificial en lógicas alternativas basadas en conjuntos borrosos, lógicas no monótonas, modales, etc.). En particular, ciertas actitudes proposicionales no pueden ser abordadas en términos de valores de verdad (por ejemplo, una actitud proposicional de duda). En consecuencia, las "proposiciones psicológicas" no deben incorporar las propiedades, la potencia o las limitaciones del cálculo de predicados ni, en general, deben compartir necesariamente las propiedades del cálculo lógico.

Esto implica poner en cuestión la equivalencia entre la dinámica de los códigos proposicionales propios del sistema cognitivo humano y la máquina de Turing y, en última instancia muestra, de acuerdo con Rivière (1986), que "la utilización de códigos proposicionales no tiene por qué implicar una concepción logicista del funcionamiento cognitivo" (pág. 17). Sin embargo, como comentaremos en el capítulo 2, en la práctica resulta difícil relajar las fuertes presuposiciones de la lógica formal en los modelos psicológicos, lo cual conducirá frecuentemente, como señalábamos unas líneas más atrás, a postular un dudoso isomorfismo entre operaciones mentales y operaciones lógicas.

El aspecto estructural de las proposiciones se refiere a la existencia de un conjunto explícito de reglas para determinar lo que constituye una proposición correctamente formada.

Aquí se pone claramente de manifiesto la raíz común que subyace al enfoque que estamos estudiando y a la gramática generativa chomskyana. En ambos casos se trata de definir un conjunto de reglas sintácticas que determinen la "gramaticalidad" de los "enunciados" obtenidos, aunque normalmente los modelos proposicionales en psicología cognitiva son mucho menos precisos y detallados, en cuanto a su "gramática" subyacente, que la formulación chomskyana. De hecho, esta "gramática de las proposiciones" debe incluir reglas de formación mucho más complejas que la relación "siguiente", utilizada para la formación de cadenas de símbolos, o que los conocidos mecanismos de transición markovianos.

Aunque, como acabamos de exponer, la proposición abstracta es el instrumento conceptual básico que usan los teóricos computacionales en su análisis de las representaciones, debe señalarse que esto no equivale a decir que las representaciones sean en sí mismas de carácter puramente abstracto. Por el contrario, parece evidente que las representaciones, en tanto que contenido de los procesos mentales, deben referirse a realidades (con todas las salvedades que el término requiere en este contexto) de tipo concreto. Por este motivo, Pylyshyn (1986) distingue entre la proposición como entidad abstracta, en el sentido que hemos venido exponiendo, y las representaciones concretas que, de acuerdo con su punto de vista, tomarían una forma análoga a la de sentencias de un "lenguaje de la mente" que permitiría expresar una misma proposición de varias formas distintas.

Al margen de las consideraciones efectuadas hasta el momento, pueden esgrimirse dos razones fundamentales que, desde el punto de vista que estamos analizando, apoyan esta idea. En primer lugar, la opacidad de la atribución de actitudes proposicionales (Fodor, 1975). Si una persona cree P, la conducta de esa persona dependerá de la forma que tome la expresión P, más que del estado de cosas al que P se refiere. Como se ha indicado repetidamente, para los teóricos computacionales hablar del contenido de las representaciones significa referirse a su contenido conceptual, y no a su presunto referente externo. El contenido de las representaciones no consiste en objetos presentes actualmente en el mundo externo; más aún, en muchos casos esos objetos ni siquiera existen, como ocurre, por ejemplo, en la actitud proposicional de "desear" algo que ahora no existe e incluso es de imposible existencia. En este contexto, la idea de opacidad en la atribución de actitudes proposicionales o, en general, de estados mentales, toma la siguiente forma: creer que "a es F" es lógicamente independiente respecto a creer que "b es F", incluso en el caso de que a sea igual a b. Si yo creo que Joan Oliver es un buen poeta, y al mismo tiempo creo que Pere Quart es un buen poeta, me estoy refiriendo al mismo estado de cosas, puesto que ambos nombres corresponden a la misma persona. Sin embargo, las creencias son distintas, en el sentido de que están asociadas a representaciones diferentes y son lógicamente independientes. Yo puedo ignorar la existencia del seudónimo del poeta en cuestión sin

que esto afecte a la creencia de que se trata (bajo su verdadero nombre) de un buen poeta. En este caso, al llegar a la biblioteca pública me dirigiré al fichero y buscaré la letra "O", mientras que si conociera el seudónimo del autor tal vez buscara en el fichero correspondiente a otra letra. Esta digresión tiene un significado preciso: partiendo de la base de que las propiedades que elicitán la conducta no se encuentran en los objetos externos, sino en la representación (o expresión) de esos objetos, es necesario señalar que las dos expresiones que estamos considerando ("Joan Oliver es un buen poeta" y "Pere Quart es un buen poeta") corresponden a una misma y única proposición, lo cual sugiere que es la forma de expresión concreta de esa proposición, y no la proposición misma, la que elicitá la conducta. En este sentido, distinguir entre dos creencias parece ser algo semejante a distinguir entre dos maneras de "decir" algo. Esto no significa, sin embargo, que tener una creencia (o cualquier otra actitud proposicional) sea equivalente a almacenar una señal de una sentencia expresada en algún lenguaje natural; una afirmación de esta índole implicaría que solamente los organismos -o los artefactos- que posean un lenguaje natural desarrollado serían capaces de poseer creencias o deseos, lo cual dejaría fuera del marco conceptual de la psicología cognitiva a los animales inferiores e incluso a los niños de corta edad. Por ese motivo pensamos que las diversas hipótesis de relatividad lingüística, así como los diversos estudios sobre, por ejemplo, los determinantes

lingüísticos en el razonamiento, no juegan en si mismas en favor del enfoque computacional sino que, en tanto que efectos "modales", constituyen algo que este enfoque debe explicar.

La posición de los teóricos computacionales es aproximadamente la siguiente: existe un lenguaje interno, de carácter conceptual (el "lenguaje de la mente"), de forma que la adscripción de una creencia a un sujeto equivaldría a afirmar la existencia (en lo que Pylyshyn llama, de forma imprecisa, "almacén activo de creencias") de una señal (inscripción) de una sentencia expresada en ese lenguaje mental. En otras palabras, las dos creencias que hemos mencionado corresponderían a dos series distintas de símbolos y, como tales, podrían formar parte de operaciones diferentes de razonamiento o acción. Naturalmente, si todo lo anterior es cierto el vocabulario del lenguaje natural propio del sujeto no debe corresponder necesariamente al vocabulario del "lenguaje mental": es perfectamente posible que existan cadenas de símbolos mentales que no tengan su expresión en el idioma del sujeto. De lo contrario podría llegarse a la identificación de la proposicional con lo verbal (que, como señala Anderson, 1978, es básicamente errónea) y entonces sería pertinente acudir a teorías que, como la de Paivio (1971), distinguen entre un código figurativo y un código verbal, sin referencia a un nivel proposicional abstracto.

La segunda razón que aboga en favor de un lenguaje específico de la mente es expuesta por Pylyshyn (1986) y radica en



la importancia de la racionalidad en la cognición. De acuerdo con Pylyshyn, la típica estructura predicativa de las sentencias es el único medio conocido para la codificación de asertos con valor de verdad de forma que pueden realizarse sobre ellos algún tipo de cálculo y ciertas operaciones propias de las teorías semánticas de tipo tarskiano. Dicho de otro modo, esta estructura predicativa garantiza la posibilidad de utilizar funciones computables cuya acción se ejercería sobre las representaciones y que, al propio tiempo, preservarían ciertas propiedades de tipo semántico.

Debe señalarse que las redes semánticas propuestas por los teóricos de la inteligencia artificial y transplantadas posteriormente a numerosos modelos propios de la psicología cognitiva no substituyen a las estructuras sentenciales propias de la lógica de primer orden; en realidad, esas redes semánticas simplemente implementan algunos aspectos de esa lógica que son especialmente adecuadas para explotar de forma eficiente las características de los actuales ordenadores (Pylyshyn, 1986). Aunque volveremos sobre la cuestión de las redes semánticas en el siguiente capítulo, debe señalarse que este tipo de construcciones constituye la aproximación básica de los modelos computacionales a la cuestión de la representación del significado. La posibilidad de definir diversos niveles de abstracción de las proposiciones permite distinguir los distintos modelos de redes semánticas (Quillian, 1969; Clark, 1969; Anderson y Bower, 1973; Kintsch, 1974; Lindsay, Norman y Rumelhart, 1975; Anderson,

1976, etc.), en función de los primitivos semánticos que adoptan (aunque, por supuesto, existen otros criterios, que serán tratados en su momento).

#### 1.4 Modelos basados en sistemas de producción

Las nociones de producción y de sistema de producción, introducidas originalmente por el grupo de investigadores de la Universidad de Carnegie-Mellon (Newell y Simon, 1972; Newell, 1973; Simon, 1975; Anzai y Simon, 1979, etc.), juegan un papel importante en el marco de los modelos computacionales, puesto que constituyen una forma de concreción del conocimiento procedimental y una aproximación, desde el enfoque computacional, a los procesos que supuestamente controlan la secuencia de fases de procesamiento de información. Por citar un ejemplo muy conocido, Anderson (1976, 1983), al presentar su teoría proposicional ACT, sugiere que el conocimiento de tipo procedimental debe representarse en forma de producciones en la memoria a largo plazo.

La distinción entre conocimiento declarativo y conocimiento procedimental proviene en gran parte, como muchos de los elementos de la psicología cognitiva, de la problemática planteada en el campo de la inteligencia artificial (Bobrow y Collins, 1975), aunque pueden encontrarse algunas ideas semejantes en otros contextos. Por ejemplo, Pascual-Leone (1976) plantea una diferenciación comparable en su

desarrollo de la distinción piagetiana entre aspectos figurativos, operativos y ejecutivos del pensamiento. Los aspectos figurativos corresponderían, en este contexto, al componente declarativo del conocimiento, mientras que los aspectos operacionales encajarían perfectamente en el concepto de conocimiento procedimental.

De hecho, la idea de sistema de producción representa la forma básica en que la psicología computacional se enfrenta al viejo problema, expuesto ya por William James, de explicar cómo ciertas condiciones contextuales o señales antecedentes elicitán la ejecución de determinadas acciones. En la clásica psicología E-R los eventos antecedentes son identificados unívocamente como estímulos, mientras que las acciones consisten en respuestas más o menos complejas evocadas por las condiciones estimulares. Como señala Allport (1980), si abandonamos la restricción de que los elementos del par condiciones antecedentes-acciones sean sucesos directamente observables este sencillo esquema asociativo se muestra altamente poderoso. Newell y Simon (1972) fueron los primeros autores que intentaron desarrollar estas potencialidades introduciendo la idea de sistema de producción como par condición-acción cuyos elementos no son, normalmente, eventos directamente observables, sino expresiones simbólicas u operadores (de Vega, 1984). De esta forma, lo que era una conexión asociativa entre estímulos y respuestas pasa a constituirse en regla procedimental que regula la aplicación de ciertas secuencias de operadores ante la presencia de

determinados estados del sistema computacional. Naturalmente, el hecho de que se trate de una regla garantiza la posibilidad de representarla en un formato proposicional y su computabilidad. En cualquier caso, Anderson (1976) ha demostrado la equivalencia formal entre sistemas de producción y máquina de Turing. Un sistema de producción consta esencialmente de dos componentes: un conjunto de reglas condición-acción (producciones) que pueden tomar la forma de instrucciones de programación del tipo IF - THEN y que poseen un carácter modular (Anderson, 1976); y una base de datos que representa el conocimiento activo que posee el sistema (Morris y Hampson, 1983). Utilizando una analogía aclaratoria, el sistema de producción puede ser visto como un conjunto de "burócratas" o expertos en cometidos más o menos específicos que están leyendo continuamente la información contenida en la base de datos a la espera de que se den las condiciones que determinarán su entrada en acción. Dada esta organización altamente modular, la acción global del sistema depende de la "cooperación" de diversos sub-sistemas, ninguno de los cuales posee un papel ejecutivo o dominante (Allport, 1980). Es importante señalar que esta última característica permite evitar el recurso a un sistema ejecutivo central peligrosamente cercano a una concepción homuncular. Sin embargo, aunque los sistemas de producción parecen estar más próximos a los modelos no jerárquicos, basados en la existencia de componentes operacionalmente distintos e interactuantes (Rowe, 1985), que a ciertas

teorías jerárquicas de la inteligencia que recurren a niveles de control metacognitivos (Sternberg, 1979, 1981), resulta muy difícil, como se verá más adelante en este mismo apartado, evitar la referencia a algunos "meta-principios" necesarios, por ejemplo, para la resolución de ciertos conflictos entre producciones.

Uno de los problemas que subyacen a este planteamiento es la forma concreta que toma la "cooperación" entre subsistemas y la relación entre producciones. En ciertas tareas relativamente simples no resulta difícil, al menos en principio, teorizar sobre cómo se produce esta relación. Por ejemplo, en el sistema de producción ideado por Newell (1973) para modelizar los procesos subyacentes a la tarea de memoria ideada por Sternberg (1966) la interacción entre producciones se produce a nivel de una memoria a corto plazo organizada en forma de "chunks". Esta memoria a corto plazo constituiría, de acuerdo con Newell, la base de datos a la que se remitirían todos los sistemas de producción implicados en esa tarea. Muchos otros modelos han utilizado también la idea de una memoria a corto plazo accesible para muy diversos módulos y que canalizaría, de alguna forma, los "mensajes" internos entre los diversos sistemas (Allport, 1980); más aún, la noción de memoria operativa como "lugar" privilegiado en el cual se concentraría gran parte de la actividad mental controlada es propia de concepciones teóricas tan importantes como la dicotomía procesos automáticos-controlados propuesta por Shiffrin y Schneider



(1977). Permítasenos señalar aquí, no obstante, que la concepción de Shiffrin y Schneider ha sido sometida últimamente a críticas bastante importantes (Cheng, 1985). En el ámbito de la percepción, en cambio, se postula la comunicación entre grupos de específicos de producciones a través de diversos "registradores" o monitores que recibirían información respecto a diversas variables sin recurrir, en principio, a un elemento central semejante a la memoria operativa (Fox, 1978).

En los modelos más simples de Newell (1973) existe una pre-suposición fundamental referida al control de la actuación de las diversas producciones. La idea básica es que el tiempo necesario para "evocar" y ejecutar un par condición-acción es independiente del número de producciones que existan en el sistema. Puede accederse a las producciones de forma ilimitada, pero estas producciones son ejecutadas de forma secuencial de acuerdo con un orden predeterminado de prioridad. Otros sistemas de producción, como en el caso del modelo ideado por Anderson (1976) permiten la ejecución en paralelo. En cualquier caso, como señala Allport (1979), para cualquier sistema distribucional es totalmente básico controlar los posibles conflictos en cuanto a demandas de procesamiento, posibilidad de interacciones ineficaces o erróneas, prioridad en la ejecución cuando existen varias producciones cuya condición inicial se cumple, etc. En general, los primeros modelos se limitaban a dejar que el orden de las producciones, tal y como aparece en el listado

del programa de ordenador, determinara su orden de evaluación. Naturalmente, este principio de resolución está implícito en cualquier esquema que evalúe secuencialmente las producciones (Klahr y Wallace, 1976). Una clase distinta de procedimientos se basa en la estructura lógica de los elementos-condición de las producciones "verdaderas". De acuerdo con esta idea, antes de ejecutar cualquier producción se evaluarán todos los pares condición-acción presentes en el programa de la tarea que se está realizando. A partir de este punto puede aplicarse toda una variedad de principios generales que determinarán la ejecución de una u otra producción. Como ejemplo sencillo puede citarse el principio de prioridad de caso especial, según el cual los pares condición-acción más específicos serán los primeros en ser ejecutados (Klahr, 1980). Finalmente, algunos principios de resolución de conflictos se basan en el orden, en la memoria a corto plazo, de los elementos correspondientes a las producciones verdaderas. En general se adopta la variante más simple, que consiste en tomar como elemento de trabajo el más cercano al principio del bucle de repaso en MCP que satisfaga la condición de aplicación de la producción correspondiente. Como señalan Klahr y Wallace (1976) y Klahr (1980, 1983), estos principios de solución de conflictos suelen ser insuficientes, por lo que está justificado el hecho de que muchos sistemas de producción recurran a reglas de producción relativamente abstractas (que, a su vez, no son sino producciones) (Hayes-Roth y Lesser, 1977) o a

algunos "meta-principios" (McDermott y Forgy, 1978) para determinar el orden de prioridad en la ejecución. Un problema especialmente complejo radica en la forma en que se produce la adquisición de nuevos pares condición-acción y, en particular, de reglas procedimentales abstractas, es decir, aplicables en contextos más amplios que los propios de los sub-sistemas más modulares. No es nuestra intención abordar exhaustivamente esta cuestión, puesto que constituye uno de los planteamientos centrales de la psicología cognitiva y, como tal, queda en buena parte fuera de los límites de este trabajo. No obstante, si parece pertinente señalar que, al margen de los enfoques clásicos del problema (por ejemplo el de la escuela de Ginebra, con todas sus extensiones y matizaciones, o el enfoque neosociacionista, Wilson, 1980), éste ha sido también abordado desde el marco propio de los sistemas de producción. En términos globales, puede considerarse como central la idea de que la adquisición de tales reglas debe ser el resultado de la adición de reglas de producción parcialmente redundantes. La primera versión de esta idea se debe a Waterman (1974) quien, en su intento de construir sistemas de producción automodificables, propuso la inclusión de una subrutina de programa cuya función era la de reunir producciones procedentes de elementos de la MCP etiquetados como condiciones, codificarlas en forma de acciones y añadir la producción creada a un "almacén de producciones". Posteriormente se han propuesto algunos sistemas en los cuales las producciones construidas intentan

emparejarse con elementos de la MCP de la misma forma en que lo hacen las demás producciones, con la salvedad de que la información obtenida de este modo se utiliza para modificar la memoria a largo plazo (en concreto, el sistema de producción) y no la propia MCP (Rychener y Newell, 1978). Este último mecanismo ofrece una explicación del paso de información de MCP a MLP en el propio marco de los sistemas de producción (Klahr, 1983). No obstante lo expuesto, no parece claro que este tipo de mecanismos pueda dar cuenta del proceso de abstracción que subyace al aprendizaje de reglas procedimentales no específicas. Como se indicará en el capítulo 4, sospechamos que el recurso a modelos distribucionales con componentes elementales no simbólicos permitirá un avance substancial en esta problemática. Resumiendo las ideas esenciales respecto a los sistemas de producción, pueden citarse algunas razones básicas que apoyan su interés y relevancia como formalismos para el conocimiento procedimental y, en términos más generales, para una aproximación computacional a las habilidades subyacentes a algunos comportamientos (Anderson, 1980).

1. El sistema está completamente "dirigido por los datos". Es decir, solamente la presencia de ciertas cadenas de símbolos en el "espacio de trabajo" del sistema (base de datos) es capaz de elicitar la ejecución de las correspondientes producciones. Una vez más, este punto muestra claramente que los sistemas de producción se enmarcan

perfectamente en el enfoque computacional de la psicología cognitiva, puesto que tanto la base de datos como los pares condición-acción están formados por cadenas de símbolos entre las que se establecen relaciones puramente sintácticas. (Kolers y Smythe, 1984). Por otra parte, el hecho de identificar el espacio de trabajo del sistema con una memoria operativa de capacidad limitada introduce un conjunto de restricciones en el procesamiento que justifican el recurso a ciertos procesos mnémicos, como la formación de "chunks", cuyo estudio es ya clásico en el ámbito de la psicología. Además, el concepto de memoria operativa es un componente importante de muchos modelos de la atención y de formulaciones teóricas bastante generales como, por ejemplo, y tal como hemos citado anteriormente, la distinción entre procesos automáticos y procesos controlados (Shiffrin y Schneider, 1977).

2. Dado que los sistemas de producción están dirigidos por los datos, siempre que este tipo de sistemas son utilizados como aproximaciones a los procesos cognitivos, deben modelizar de forma explícita la estructura de los procesos de control. Esto ocurre, por ejemplo, en el análisis de Newell (1973) sobre la tarea de memoria de Sternberg que, además, es generalizable al paradigma experimental de Posner, a tareas de aritmética mental, etc. (Chase, 1975; Morris y Hampson, 1983). El carácter homogéneo, distribucional y modular de los sistemas de producción



garantiza la inexistencia de sistemas ocultos de control, por lo que dicho control será el resultado de la existencia de las cadenas de símbolos adecuadas en la base de datos. Podremos decir entonces que estas cadenas específicas de símbolos identifican objetivos. De esta forma, podremos hablar de una definición computacional del concepto de objetivo o meta; además, puede afirmarse que la base de datos incluye en todo momento la totalidad de los contenidos que ocupan la atención del sistema, incluyendo los objetivos activos del procesamiento. En la medida en que este espacio de trabajo o base de datos se identifique con la memoria operativa del sistema, y ésta a su vez con el contenido de la conciencia, podrá hablarse de un mayor o menor nivel de acceso consciente a los procesos cognitivos. La toma de postura en cuanto a esta cuestión tiene repercusiones fundamentales no solamente a nivel teórico sino en el ámbito metodológico (Ericsson y Simon, 1980, 1981, 1984; Froufe, 1985). Generalizando lo que acabamos de exponer en este segundo punto, puede afirmarse que, desde un punto de vista computacional, los sistemas de producción permiten la integración de principios psicológicos generales (Klahr, 1980), en el sentido de que incorporan una teoría de la estructura de control y de la representación que subyace a una amplia gama de habilidades cognitivas del sujeto. Como señaló en su momento Newell (1973), el sistema de producción se convierte en portador de los supuestos psicológicos básicos, es decir, la estructura del sistema de producción es considerada como

la estructura del sistema humano de procesamiento de información (pág. 516).

3. Los sistemas de producción son altamente modulares; por tanto, pueden ser extendidos uniformemente sin que se produzcan alteraciones importantes en el sistema. Además, las producciones individuales pueden ser componentes significativos del sistema global, de forma que el tratamiento en términos de creencias u objetivos, tal y como sugiere el enfoque computacional, se hace más verosímil (Pylyshyn, 1986).

Aunque los modelos basados en sistemas de producción han sido propuestos en muy diversos ámbitos de trabajo (Anderson, 1980), probablemente uno de los campos de acción en el que han conocido un mayor desarrollo es el de la solución de problemas. A partir del trabajo pionero de Newell y Simon (1972), puede afirmarse que una parte muy substancial de los modelos propuestos para la solución de problemas en el marco del enfoque computacional se basan en sistemas de producción. Más aún, la introducción de este tipo de sistemas por parte del grupo de Carnegie-Mellon se ha realizado en primera instancia como un intento de aproximarse a la solución de problemas en sujetos humanos. Analizaremos brevemente algunas características de este enfoque de los procesos de solución, puesto que, a nuestro entender, los trabajos realizados en este ámbito ilustran algunos aspectos

importantes de la aproximación computacional a lo que, en términos generales, denominaremos el "cambio". Como señalaremos en el apartado 2.3, la cuestión de las transformaciones que sufre el sistema cognitivo (y su correspondiente traducción conductual), ya sea por diversos tipos de aprendizaje, por medio de lo que se ha denominado a veces "creatividad" o por procesos de cualquier otra índole, resulta especialmente problemática para este tipo de enfoque.

Como es sabido, la aproximación computacional-discreta a la solución de problemas se basa en unas pocas ideas esenciales: En primer lugar, la estructura del ambiente de la tarea, que define el conjunto de operaciones lícitas y la meta de la tarea a realizar. La estructura de la tarea interactúa con los límites de la memoria a corto plazo, puesto que las diversas "trayectorias de solución" pueden poseer exigencias de memoria distintas. De hecho, la estructura de la tarea no puede, en general, ser aprehendida completamente por el sujeto, el cual se representa una parte de esta estructura en lo que constituye su espacio-problema propio (Newell y Simon, 1972; Anzai y Simon, 1979). Este espacio-problema toma la forma de una red cuyos nodos corresponden a los diversos estados de conocimiento que puede alcanzar el sujeto. Podrá distinguirse entonces entre lo que Greeno (1978) denomina "conocimiento estructural", que se referiría a la aprehensión de la estructura global de la tarea, y el conocimiento parcial o incompleto, que es postulado por la mayoría

de los autores que trabajan en esta línea (Anzai y Simon, 1979).

A partir de cualquier estado de conocimiento, el sujeto podrá escoger -siempre de acuerdo con un conjunto de reglas condicionales- entre un conjunto de operadores disponibles aquel que sea más pertinente. Muchos modelos incorporan en este punto un conjunto de procesos evaluativos basados en la determinación de la distancia desde el estado actual hasta la solución. Al propio tiempo, un mecanismo de planificación proporcionaría una anticipación de algunos estados futuros a los que conduciría la aplicación de cierto operador. La combinación de procesos de planificación y evaluación permite anticipar y valorar las consecuencias de escoger una u otra trayectoria dentro del árbol de soluciones.

La teoría general asume la distinción entre algoritmos y heurísticos. Estos últimos se distinguirían por el hecho de que no conducen de forma necesaria y segura a una solución; su función sería más bien la de reducir de forma radical el tamaño del espacio-problema (por ejemplo, mediante la formulación de sub-objetivos) o anticipar y evaluar algunas líneas de solución (heurísticos de planificación y de análisis medios-fin). Debe señalarse, no obstante, que la distinción que hemos expuesto no resulta fundamental, puesto que, desde el planteamiento que estamos analizando, los heurísticos deben adoptar necesariamente una forma algorítmica. Desde la perspectiva computacional-discreta, la actividad básica del sistema cognitivo es la ejecución de operaciones de

procesamiento consistentes en la manipulación secuencial y paso a paso de cadenas de símbolos, y este tipo de operaciones son expresables siempre en forma de programas ejecutables por un ordenador digital y, en última instancia, como conjuntos de algoritmos.

En cualquier caso, los sistemas de producción constituyen una aproximación poderosa a la solución de problemas. Resulta claro que un modelo basado en reglas procedimentales de tipo condicional proporciona una aproximación unificada al conjunto de procesos que acabamos de delinear rápidamente. Una buena prueba de ello lo constituye su aplicación a una amplia variedad de tareas; pueden citarse, a este respecto, la utilización de sistemas de producción para modelizar la ejecución de sujetos humanos en problemas de criptoaritmética y otras tareas aritméticas (Newell y Simon, 1972; Reed, 1984; Reed, Dempster y Ettinger, 1985; Sweller, Mawer y Ward, 1983), problemas tipo "rompecabezas" o de "movimientos" (jarras de agua, torre de Hanoi) (Atwood y Polson, 1976; Jeffries, Polson, Razran y Atwood, 1977; Atwood, Masson y Polson, 1980; Anzai y Simon, 1979; Karat, 1982; Kotovsky, Hayes y Simon, 1985), razonamiento silogístico (Fisher, 1981), tareas de "razonamiento científico" (Good, 1984; Gorman, 1986), problemas de ajedrez (Newell y Simon, 1972), etc.

Parece bastante claro que la amplia utilización de los sistemas de producción en la modelización de la solución de problemas se justifica por el carácter procedimental que



estos sistemas incorporan. En este sentido, como señalan Newell y Simon (1972), este tipo de sistemas resulta especialmente adecuado para aproximarse a tareas en las cuales se exige un amplio repertorio de respuestas ante situaciones que pueden ser bastante diversas. Es importante señalar, en todo caso, que al margen de un pequeño conjunto de reglas de tipo general (que corresponden a la serie básica de heurísticos), este tipo de sistemas introduce un gran número de reglas de cómputo de carácter local, específicas de tarea. La revisión de los trabajos realizados en este ámbito pone también de manifiesto que la derivación de nuevas reglas de producción (incluyendo especialmente las de carácter relativamente general) solamente puede realizarse, en el marco de esta aproximación, por aplicación de reglas condicionales anteriormente existentes. Como veremos en el siguiente capítulo, esto plantea una dificultad muy considerable a los modelos basados en sistemas de producción.

Señalaremos finalmente que, de acuerdo con Anderson (1976), los sistemas de producción, dado su carácter procedimental, garantizarían la reactividad del sistema ante la información externa, evitando de alguna forma el solipsismo propio del formalismo proposicional puro adoptado por otros autores. En el siguiente capítulo volveremos brevemente sobre esta cuestión.

2. Símbolos y procesos : Bases para un análisis crítico de los modelos computacionales discretos
-

## 2.1 Niveles de una teoría general del procesamiento de información

Si se realiza una revisión, siquiera superficial, de los artículos recientes de análisis teórico en el marco de la psicología cognitiva puede constatarse la existencia de frecuentes referencias al libro de D. Marr "Vision" (Marr, 1982). Este énfasis sobre el trabajo de Marr es debido fundamentalmente, y al margen de la indiscutible influencia de su análisis de la visión humana (Roth y Frisby, 1986), a su aportación en el ámbito de la sistematización de los niveles que debe poseer una teoría coherente y general del procesamiento de información. El análisis de estos niveles es fundamental, puesto que con frecuencia existen serias dudas sobre la comparabilidad de ciertos modelos entre sí, dada la posibilidad de que pertenezcan a niveles distintos; por otra parte, la pertenencia de un modelo a uno u otro nivel tiene implicaciones importantes para su contrastabilidad empírica. En términos generales, una visión coherente de los niveles que debe poseer una teoría del procesamiento de información es, indudablemente, un instrumento conceptual necesario para el análisis y evaluación de la gran cantidad de modelos propuestos en el ámbito de la psicología cognitiva. La utilidad de esta sistematización se pondrá de manifiesto, al margen de las consideraciones realizadas en este apartado, al tratar diversos aspectos de este trabajo (en particular, en la polémica sobre el nivel que ocupan los modelos

distribucionales de memoria: McClelland y Rumelhart, 1985; Broadbent, 1985; Rumelhart y McClelland, 1985).

De acuerdo con Marr (1982), pueden distinguirse tres niveles fundamentales en una teoría del procesamiento de información:

1. **Teoría de la computación.** Intentaría responder a preguntas del tipo: ¿cual es el objetivo de la computación?, ¿cual es la lógica de la estrategia utilizada, y por qué es apropiada?. La teoría de la computación parte de una especie de caja negra o módulo de computación que ejecuta una cierta actividad y especifica el input de esta computación y el propósito de la misma; obviamente, tanto la naturaleza del input como el propósito de la computación introducen restricciones sobre qué computación debe ser ejecutada. Consideremos, por ejemplo, una computación que detecte los bordes de los objetos en base a la intensidad de la luz reflejada sobre el objeto y su entorno. Marr señala que el input sería, en este caso, un "campo de luz" bidimensional en el cual cada punto tendría asociada una cierta intensidad luminica. El propósito de la computación es el de detectar los lugares en que se produzcan cambios de intensidad muy bruscos, puesto que éstos indicarian con toda probabilidad la existencia de un borde del objeto. Tanto la naturaleza del input como el objetivo de la computación sugieren la ejecución de una secuencia de procesamiento muy concreta: localizar puntos contiguos en los que la segunda derivada de

la función que relaciona posición e intensidad cambie de signo.

Como referencia aclaratoria, el propio Marr señala que, en el campo de la lingüística, el modelo de competencia de Chomsky (1965) encajaría plenamente en el terreno de la teoría de la computación, mientras que el análisis de la "performance" o ejecución sería propio del segundo nivel de descripción (nivel algorítmico). En el nivel de la teoría computacional no posee ningún interés el hecho de que la teoría sea construida en forma de programa para un ordenador, como un conjunto de axiomas o como una serie de reglas; no interesa tampoco cómo es representada la información ni la secuencia de transformaciones a las cuales es sometida esta representación. La cuestión es qué función debe ser computada, no cómo se realiza esa computación.

Debemos señalar, en primer lugar, que el propio Marr (1982) señala explícitamente que el nivel de la teoría computacional no es el nivel propio de una teoría psicológica. Como veremos algunas líneas más adelante, no es difícil aventurar algunas razones que avalan esta afirmación.

**2. Nivel algorítmico.** Las preguntas básicas que se plantearían en este nivel son del tipo: ¿cómo puede implementarse una teoría computacional concreta?; ¿de qué forma son representados el input y el output, y cual es el algoritmo que permite realizar la transformación de uno en otro?. Se trata, en consecuencia, de especificar cómo se realiza una determinada computación cuya necesidad ha quedado



establecida en el primer nivel, es decir, cual es el algoritmo que describe la secuencia de operaciones de procesamiento que han permitido pasar del input al output. En el ejemplo que se mencionaba anteriormente, el algoritmo deberá incorporar alguna forma de representación de las intensidades luminicas, y será necesario utilizar operaciones específicas para localizar y organizar los puntos en los que se produzcan cambios bruscos de intensidad. Marr sugiere que dichos puntos (cuya localización se realizará por el procedimiento explicado anteriormente) pueden ser conectados por líneas, de forma que éstas representarían, hipotéticamente, los bordes del objeto percibido. Los resultados de esta operación para las distintas bandas de frecuencia serían comparadas, y aquella configuración que apareciera en todas las bandas sería tomada como la forma del objeto percibido.

La distinción entre teorías de la computación y algoritmos es reflejo de un hecho evidente, a saber, que cualquier función input-output computable puede ser realizada por muy diversos procedimientos efectivos. De hecho, técnicamente hablando, puede especificarse un número infinito de posibles procedimientos para el cálculo de cualquier función computable (Rogers, 1967), aunque, naturalmente, algunas de las variaciones implicadas pueden considerarse como puramente triviales. Incluso en el caso de que conociéramos exactamente cual es la función ejecutada en un proceso cognitivo dado, puede ser problemático identificar el procedimiento usado para computarla (Marr, 1977). Todo ello sugiere la necesidad

de introducir algún tipo de criterio valorativo de la amplia variedad de algoritmos susceptibles de ejecutar una función determinada (Johnson-Laird, 1983).

3. **Implementación física.** Resulta absolutamente trivial afirmar que cualquier algoritmo que ejecute una función computable dada exige un substrato material en el cual se desarrollen las operaciones físicas que permiten la realización efectiva de la computación. Nos limitaremos, de momento, a señalar que la posición funcionalista propia de los modelos computacionales conduce a ignorar una posible fuente de restricciones al enorme poder de formalización de este tipo de aproximación. Como señalaremos más adelante, la posición funcionalista es perfectamente sostenible siempre que se reconozca el carácter de teoría de la competencia que, a nuestro entender, poseen los modelos de base proposicional. En caso contrario, pensamos que la arquitectura material del sistema debe introducir restricciones empíricas que no pueden ser ignoradas.

Pensamos que la aproximación propuesta por Marr es altamente relevante para el análisis del estatus de los modelos computacionales en psicología. El problema básico podría plantearse en los siguientes términos:

Como se indicó repetidamente en el capítulo 1, los modelos computacionales en psicología encuentran buena parte de sus raíces en el marco de la tradición formalista desarrollada en las matemáticas y la lógica. La pretensión de los lógicos al desarrollar el cálculo proposicional era el de obtener

una sintaxis para la representación de los argumentos lógicos. En su versión más evolucionada (que incluye variables y cuantificadores), el cálculo de predicados constituye un modelo adecuado de esa sintaxis y permite representar teoremas y demostraciones de carácter general. Es cada vez más claro que el cálculo de predicados es suficiente, como formalismo subyacente a la computación, en el mismo sentido en que lo es la máquina de Turing: cualquiera de los dos instrumentos conceptuales puede representar cualquier función computable dada. Dada la característica de equipotencia, lo único que diferencia al cálculo de predicados de los instrumentos de computación del tipo máquina de Turing es el hecho de que estos últimos prescriben la forma concreta en que se realiza la computación; es decir, una máquina de Turing (o un ordenador) introducen limitaciones explícitas a la forma en que se realiza la computación, mientras que el cálculo de predicados se ve libre de tales restricciones (Anderson, 1978). Como señalaron en su momento los teóricos de las "ciencias de la computación" cualquier proceso exhaustivamente descrito puede ser convertido en un procedimiento que opera sobre datos estructurados proposicionalmente. En consecuencia, las descripciones proposicionales en si mismas no constituyen modelos, sino que simplemente reexpresan las observaciones en una sintaxis formal (Hayes-Roth, 1979). Igualmente, el hecho de que un sistema pueda ser descrito en términos de símbolos y de reglas definidas sobre esos símbolos no significa que el sistema contenga realmente un

sistema simbólico (Kosslyn y Hatfield, 1984) del mismo modo que, como han puesto de manifiesto repetidamente los filósofos, es muy distinto hablar de procesos descritos por reglas que referirse a procesos gobernados o dirigidos por reglas. Volveremos sobre esta cuestión en el siguiente apartado. Lo que interesa resaltar aquí, de momento, es que si, como se ha señalado anteriormente, el cálculo de predicados (con algunas extensiones) posee la característica de suficiencia en tanto que formalización de funciones computables, si al mismo tiempo no introduce restricciones sobre la forma concreta en que puedan realizarse dichas funciones, y si aceptamos el carácter computacional de los procesos cognitivos, queda perfectamente claro que poseemos evidencia en favor de dos ideas fundamentales: en primer lugar, es esencial imponer restricciones teóricas a la forma que toman las representaciones y a los procesos relacionados con ellas; en segundo término (pero no menos importante), los modelos computacionales poseen, al menos en su formulación presente, el carácter de modelos de la competencia y, en la terminología empleada por Marr, se enmarcan en el nivel de teorías de la computación.

Debe señalarse que algunos teóricos proposicionalistas han afirmado explícitamente que sus modelos no incluyen todas las potencialidades del cálculo de predicados (Kintsch, 1974; Schank, 1979, 1982), lo cual introduciría, al menos potencialmente, algunas restricciones sobre el poder computacional que dichos modelos podrían poseer. Por ejemplo,

Schank (1979) rechaza la utilidad del cálculo de predicados como modelo psicológico y, en particular, se opone a la idea de que los procesos cognitivos humanos se basen en procedimientos de prueba de teoremas. Este último punto es importante, puesto que con frecuencia se ha postulado la existencia de una especial afinidad entre las representaciones proposicionales y tareas como la prueba de teoremas y, más en general, la elaboración de inferencias (Pylyshyn, 1973; Fodor, 1975; Anderson, 1976; Anderson y Kline, 1977). De hecho, un argumento importante para la introducción del formalismo proposicional como supuesto formato de las representaciones radica en el hecho de que, al incorporar valores de verdad abstractos, permiten reducir la complejidad y el número de las inferencias que deben realizarse en una tarea concreta (Anderson, 1978). Por ejemplo, en un sistema proposicional abstracto no es necesario postular la existencia de reglas de inferencia distintas para obtener conclusiones a partir de frases activas y pasivas. Sin embargo, como se indicó ya en el capítulo precedente, el concepto psicológico de proposición debe ser lo bastante flexible como para incorporar enunciados sobre los cuales no tendría sentido postular un valor de verdad; de ahí que la incorporación de las potencialidades del cálculo de predicados no sea una exigencia ineludible en la construcción de modelos proposicionales.

Uno de los problemas de los modelos proposicionales que rechazan explícitamente algunas de las potencialidades del



cálculo de predicados es la falta de definición completa de la gramática formal de sus representaciones. Por ejemplo Kintsch (1976/1979), en sus trabajos sobre memoria para prosa, no entra en profundidad en el análisis de las posibles reglas de formación de sus "textos básicos"; contrariamente, se limita a señalar que tales textos se definen como conjuntos o series de proposiciones susceptibles de representar el significado de cierto número de textos específicos, y a sugerir la existencia de ciertos procesos de superposición de argumentos que servirían para conectar esas proposiciones. El propio autor reconoce que se encuentra lejos de definir totalmente las reglas sintácticas propias de su modelo. Está claro, sin embargo, que cualquier modelo proposicional debe incorporar ciertas reglas de formación (cierta gramática) que permitan determinar qué es una proposición bien formada. Además, para el psicólogo es fundamental que tales reglas de formación correspondan a lo que sucede realmente en los procesos cognitivos humanos, de manera que la gramática del modelo proposicional propuesto corresponda a las reglas de formación que regulan la construcción de representaciones en el plano cognitivo. Si no se definen estas reglas de formación, el modelo pasa automáticamente a ser no falsable. De acuerdo con Anderson (1979), un modelo de la representación que no contenga presuposiciones respecto a los procesos definidos sobre esa representación no constituye una construcción científica. Como señala muy acertadamente Rivière (1986) probablemente sea imposible demostrar la no

equivalencia entre la competencia de los sujetos humanos y las potencialidades del cálculo de predicados sin definir de forma exhaustiva las reglas de formación incluidas en cualquier supuesto modelo alternativo. En cualquier caso, y citando nuevamente a Rivière, el único modelo proposicional suficientemente desarrollado como para que su poder de computación pueda ser establecido (el modelo ACT de Anderson, 1976) posee un poder computacional equivalente al de la máquina de Turing (y, por tanto, añadimos nosotros, equivalente también al del cálculo de predicados). De esta forma, Anderson se ve abocado a explicar las obvias diferencias entre la actuación de una máquina de Turing y la de los sujetos humanos en base, sobre todo, a la existencia de capacidades de memoria desiguales y de una probabilidad de error distinta de cero en la recuperación de información, en el caso de los sujetos. Es casi innecesario añadir que nos encontramos, una vez más, ante el clásico lenguaje de las teorías de la competencia y que, por tanto, los problemas típicos de estas teorías en cuanto que construcciones científicas son aplicables también a los modelos computacionales en psicología cognitiva.

Como señala de Vega (1981), resulta sencillo observar el paralelismo que existe entre los planteamientos de una teoría de la competencia y el programa propio de los lógicos idealistas contemporáneos (especialmente el programa fregeano). La idea común a todos estos planteamientos es el establecimiento de una línea divisoria entre el mundo de la

competencia (que, de alguna manera, correspondería al Pensamiento fregeano) y el de la actuación (es decir, el ámbito de los pensamientos concretos; Frege, 1918). Puede afirmarse, en líneas generales, que las teorías de la competencia son formulaciones de sistemas de reglas u operaciones que describen en un lenguaje abstracto los tipos de conocimiento que puede ejecutar idealmente la mente humana. Lo que pretenden los teóricos proposicionales es precisamente elaborar un sistema formal que permita interpretar la mayoría de los comportamientos inteligentes con la máxima economía conceptual. Resulta casi superfluo decir que cualquier desviación observada en la conducta externa respecto a las prescripciones del nivel competencial es atribuida, en el marco de estas teorías, a influencias ajenas al propio corpus de competencia (fatiga, errores de memoria, etc.) (Richelle, 1971). Podemos abordar la cuestión desde un ángulo ligeramente distinto, a saber, el de la relación entre las funciones computadas por un sistema y la arquitectura funcional del mismo. Tomemos nuevamente el ejemplo de la máquina de Turing. Aunque, como se indicó en el capítulo anterior, esta máquina es universal (en el sentido de que puede ejecutar cualquier función computable), esto no significa que lo haga exactamente igual que, por ejemplo, un ordenador digital. La complejidad de la secuencia de operaciones ejecutadas por la máquina de Turing cambia de acuerdo con la naturaleza de la tarea que debe ejecutar y, cosa mucho más importante, esta dependencia respecto a la tarea es distinta a la que pueda

establecerse en el caso de otro tipo de máquinas. Por ejemplo, el número de pasos exigidos para que una máquina de Turing lea un cierto número de símbolos presenta una relación cuadrática con la cantidad de símbolos implicada. En cambio, en una "máquina de registro", una arquitectura que posee lo que normalmente se denomina "acceso aleatorio" a la memoria del sistema puede, bajo ciertas condiciones, no existir relación entre número de símbolos y tiempo de lectura. En consecuencia, una arquitectura de este último tipo puede ejecutar directamente ciertos algoritmos que, en cambio, no pueden ser llevados a cabo por la máquina de Turing. Por este motivo podremos afirmar que la equivalencia entre las dos arquitecturas es débil, puesto que, aunque ambas puedan computar las mismas funciones, no lo hacen ejecutando los mismos algoritmos.

Naturalmente, la máquina de Turing puede emular la secuencia de estados de la máquina de registro mediante un proceso que comprendería dos fases. En primer lugar, la máquina de Turing debe ser preparada para computar las funciones realizadas por la máquina de registro; la diferencia radicará en el hecho de que una operación primitiva para la máquina de registro será ejecutada por la máquina de Turing mediante una secuencia de operaciones. Esto equivale a emular la arquitectura funcional de la máquina de registro. En segundo lugar, lógicamente, se ejecutará el algoritmo resultante, de forma que la máquina de Turing realizará la tarea propuesta simulando la ejecución realizada por la máquina de registro.

La distinción entre la realización de una computación por ejecución directa de cierto algoritmo por cierta arquitectura funcional (que posee ciertas operaciones primitivas), o por simulación de esa arquitectura en un sistema distinto (con operaciones primitivas diferentes) posee implicaciones importantes. Por una parte, subyace a una de las más conocidas críticas al enfoque computacional: la confusión, difícilmente evitable, entre presuposiciones teóricas de cierto modelo computacional y necesidades de programación del sistema (Johnson-Laird y Wason, 1977; de Vega, 1984). Parece bastante claro que los modelos computacionales, cuando "descienden" desde el nivel de teorías de la computación al nivel algorítmico, deben incorporar necesariamente algunas presuposiciones relacionadas con la arquitectura funcional en la cual se desee implementar el modelo. Y, más importante aún, esas características de la arquitectura funcional poseen consecuencias empíricas irreductibles. Retomando el ejemplo anterior, una máquina de Turing que emule la arquitectura funcional de una máquina de registro seguirá mostrando, a diferencia de ésta, una relación cuadrática entre número de ítems a leer y tiempo de lectura (Pylyshyn, 1986). Puede citarse, como ejemplo de ámbito más general, la diferencia entre arquitecturas funcionales seriales y paralelas. A nivel de estados lógicos (es decir, a nivel de teoría de la computación), no existe diferencia entre los dos tipos de procesamiento. La teoría de autómatas ha demostrado que una máquina serial puede, en principio, computar cualquier



función ejecutada por una máquina que funcione en paralelo. Sin embargo, los algoritmos utilizados en cada caso son totalmente distintos y producen diferencias muy importantes a nivel empírico (Fahlman, 1979, 1981). Esto es cierto hasta el punto de que se está produciendo, en el ámbito de la inteligencia artificial, un importante resurgimiento del interés hacia los dispositivos en paralelo, puesto que los sistemas seriales se han mostrado muy poco eficaces (en particular, en cuanto al tiempo utilizado) en la ejecución de ciertas funciones de computación. Este resurgimiento se demuestra de muy diversas formas (y con distintas implicaciones): los "actores" de Hewitt, la "teoría de la sociedad" de Minsky y Papert (1969), los sistemas de producción distribuidos del grupo de Carnegie-Mellon (Newell y Simon, 1972; Newell, 1973, 1980), las redes semánticas en paralelo (Hinton, 1981), "redes de creencias" (Pearl, 1986), teorías de ordenación causal (de Kleer y Brown, 1986; Iwasaki y Simon, 1986a, 1986b), etc. Lo que interesa aquí, sin embargo, es el hecho de que la contrastación empírica de modelos computacionales en el ámbito de la psicología cognitiva no puede ignorar la presencia de restricciones específicas debidas a la arquitectura funcional del sistema.

Los teóricos computacionales más reflexivos, sobre todo Pylyshyn (1986) son conscientes de que la referencia a la arquitectura funcional del sistema ofrece la oportunidad (única, dado el carácter de teoría de la competencia que poseen los modelos computacionales) de reducir los grados de

libertad en la construcción de modelos formales de la cognición y proporcionar una vía de comparación entre esos modelos. De ahí su insistencia en la necesidad de definir las operaciones primitivas de un sistema computacional, las cuales estarían implementadas como operaciones elementales de su arquitectura funcional y, sobre todo, su exigencia de "equivalencia funcional fuerte" como criterio de contrastación entre modelo y realidad. Naturalmente, Pylyshyn no renuncia a la idea de que tales operaciones primitivas sean expresadas como operaciones elementales en cierta notación: por ejemplo, al proponer la realización de funciones computacionales relacionadas con la representación espacial, se reconoce como primitiva, entre otras, la relación "a la derecha de"; de acuerdo con Pylyshyn, esta relación tomaría la forma de un "predicado primitivo" que, cuando se aplica a los nombres (cadenas de símbolos que representarían objetos) evalúa la relación direccional entre ellos. En términos generales, la propuesta de Pylyshyn consiste en que las restricciones impuestas por la arquitectura funcional del sistema (básicamente, en forma de establecimiento de operaciones primitivas) son equivalentes a la introducción de propiedades formales fijas en un sistema notacional. Estas propiedades limitarían, por ejemplo, el poder representacional del sistema, de la misma forma en que una notación que incorpore los presupuestos de la geometría euclidiana no puede expresar propiedades no euclidianas, o un álgebra de Boole no puede expresar ciertas propiedades de un álgebra de De

Morgan. Globalmente, la tarea sería la de construir, una vez más, un sistema formal cuya gramática fuera equivalente a la de las operaciones cognitivas humanas y, en particular, que compartiera con éstas un cierto número de relaciones y propiedades primitivas. La idea fundamental es la de conseguir que el "nivel de agregación" de ambos sistemas sea semejante; es decir, se trataría de construir un sistema formal que ejecutara directamente, y no por emulación de la arquitectura funcional, las operaciones primitivas del intelecto humano. En esto radica, para Pylyshyn, el concepto de "equivalencia funcional fuerte", puesto que la equivalencia de operaciones y relaciones primitivas garantizaría, al menos en gran parte, la semejanza de las conductas externas (Pylyshyn, 1986, pp.107 y ss.).

Fodor (1968, 1981a) trata exhaustivamente la cuestión de la equivalencia funcional entre modelos computacionales y procesos cognitivos humanos. Una de las ideas básicas es que una condición necesaria para que el modelo formal adquiera un valor explicativo es la equivalencia entre las observaciones conductuales potenciales del sistema y del sujeto. Una idea semejante ha sido expuesta por Bower (1975), quien señala explícitamente que la semejanza entre las conductas manifiestas de un ordenador y de un sujeto humano no garantiza la equivalencia entre los procesos internos responsables de la conducta ni la posibilidad de extrapolación de esa semejanza a otras conductas potenciales. Estas observaciones muestran claramente la insuficiencia de un criterio

de equivalencia funcional débil para dotar a los modelos computacionales de valor aproximativo a las conductas de los sujetos humanos.

Un nuevo problema se plantea cuando se intentan analizar los requisitos que, supuestamente, deben permitirnos establecer la equivalencia funcional fuerte entre dos sistemas. Es factible, por ejemplo, aplicar la noción de equivalencia fuerte propuesta por Pylyshyn a fin de obtener criterios de comparación entre programas; en este caso, podría establecerse un criterio de complejidad basado en la existencia de ciertas relaciones topológicas entre dos programas al estilo, por ejemplo, del método de minimización de diagramas de flujo de Fosdick y Osterweil (1976). Esto introduce algunas cuestiones interesantes; por ejemplo, la equivalencia entre dos arquitecturas funcionales se establece en un nivel distinto (más abstracto) al de los lenguajes de programación. Dos lenguajes de programación pueden ejecutar el mismo algoritmo, representado por el correspondiente diagrama de flujo, utilizando un número de instrucciones bastante distinto; sin embargo, es bastante obvio que no interesa considerar las instrucciones como "operaciones primitivas" en el sentido que comentábamos anteriormente. No obstante, debe señalarse que es difícil, incluso, mantener la idea de que el lenguaje de programación utilizado no tenga implicaciones teóricas y empíricas importantes. Como afirma Pylyshyn (1981), la relación entre algoritmo y lenguaje de programación es semejante, de alguna forma, a la establecida entre proposición y

sentencia. Hay muchas sentencias que pueden expresar una misma proposición; de la misma forma, un mismo algoritmo puede ser expresado en varios lenguajes de programación. En el marco del enfoque computacional, una proposición es identificada con una sentencia en algún lenguaje formal (por ejemplo, el cálculo de predicados), mientras que el algoritmo es identificado con un programa en algún lenguaje canónico de programación. Lo que ocurre es que, en ambos casos, el lenguaje adoptado no puede ser cualquiera, y toda decisión a este respecto posee consecuencias teóricas y, al menos potencialmente, empíricas. No puede utilizarse cualquier lenguaje para expresar proposiciones; el lenguaje adoptado debe estar libre de ambigüedades léxicas y sintácticas y debe ser lo suficientemente rico para aprehender distinciones relevantes. De forma semejante, un lenguaje de programación debe ajustarse a las necesidades del tipo de programas a ejecutar; en particular, no todos los lenguajes son apropiados para expresar un "algoritmo cognitivo" (por ejemplo, en algunos aspectos el lenguaje LISP parece más adecuado que otros, a pesar de sus insuficiencias). En consecuencia, el programa de trabajo del enfoque computacional en psicología debe incluir explícitamente el objetivo de hallar un lenguaje de programación que exprese y ejecute los procesos cognitivos en forma semejante (y, por supuesto, hablamos de "semejanza fuerte") a los sujetos humanos. Sin embargo, en este nivel nos parece extraordinariamente difícil construir un lenguaje de programación en el cual no se produzca



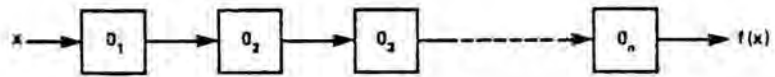
confusión entre presuposiciones teóricas y cuestiones técnicas. En rigor, y en tanto que la estructura de la arquitectura funcional de un sistema dependa, de alguna forma, de su arquitectura material, el objetivo nos parece casi inviable (este punto será tratado en este mismo apartado). Por otro lado, la distinción entre algoritmo y lenguaje de programación nos lleva a pensar, en la línea expuesta por Rumelhart y McClelland (1985) (aunque estos autores se muevan en un marco bastante distinto al que estamos exponiendo) en la existencia de sub-niveles diferenciales dentro del nivel algorítmico propuesto por Marr.

Debe señalarse, por otra parte, que los teóricos computacionales, al restringir su marco conceptual a la consideración de un tipo de procesamiento muy concreto (el procesamiento de tipo serial) han introducido también serias limitaciones tanto a los lenguajes como al "estilo" de programación al que estos lenguajes conducen. El énfasis en este tipo de lenguajes ha venido acompañada por la preeminencia de un tipo de arquitectura muy determinado. Backus (1978), en una Comunicación presentada al recibir el premio anual de la Association for Computing Machinery, (y que lleva el sintomático título de "Can programming be liberated from the von Neumann style?") señala con mucha claridad que la suposición, presente aún en algunos teóricos computacionales, de que existe un sólo tipo de computador es la base de la creencia de que existe también un único tipo de lenguaje de programación, el lenguaje convencional basado en las

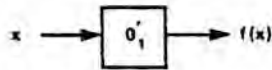
presuposiciones adoptadas por von Neumann. Esta orientación ha limitado el desarrollo de lenguajes basados en principios distintos y ha privado a los diseñadores de un fundamento intelectual para la construcción de arquitecturas distintas. Ciertamente, estas actitudes están cambiando con rapidez, como lo demuestra el desarrollo de ideas innovadoras en el diseño de computadores, incluyendo sugerencias tan interesantes como la preparación de "ordenadores biológicos", el diseño de "máquinas conectivas" en paralelo (Hillis, 1981) o de arquitecturas comunicativas paralelas (Debrunner, 1983; Pollack, 1982). Es casi superfluo añadir que el análisis de la repercusión de estas innovaciones tecnológicas en inteligencia artificial (y, en un sentido distinto, en la psicología cognitiva) es una tarea fundamental y puede hacer cambiar de modo bastante radical el enfoque actual de muchos psicólogos cognitivos. Pensamos que el desarrollo de modelos en paralelo y con procesamiento continuo es un primer indicio de este cambio, y como tal lo analizaremos en capítulos posteriores.

El criterio de complejidad permite definir la operación primitiva como aquella que utiliza siempre los mismos recursos de computación (tiempo o memoria) independientemente de las características del input. (Figura 2.1). Esta idea se relaciona con el concepto de equivalencia computacional propuesto por Simon (1978), el cual introduce la posibilidad de existencia de un factor de proporcionalidad relacionado con características no esenciales de los dispositivos de

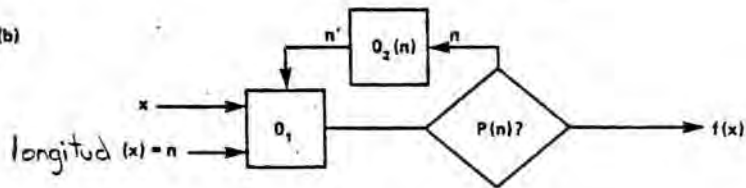
(a)



• es equivalente en complejidad a



(b)



no es equivalente en complejidad a

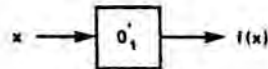


Figura 2.1 Ejemplos de algoritmos de complejidad equivalente (arriba) y no equivalente (abajo)

computación (por ejemplo, la rapidez en el funcionamiento de sus elementos). Pero, aunque este tipo de criterio (junto con otros, puesto que probablemente no es suficiente) permita establecer un cierto nivel de comparabilidad entre programas, queda sin resolver la cuestión de la equivalencia de cualquiera de esos programas con el funcionamiento de los procesos cognitivos humanos. En el caso de los sujetos humanos se hace necesario acudir a medidas conductuales de diverso tipo, lo cual plantea de inmediato un conjunto de problemas metodológicos relacionados, en gran parte, con la utilización y análisis de protocolos verbales. Aunque no entraremos en el análisis de esta cuestión, dada la existencia de abundante literatura sobre el tema (Newell y Simon, 1972; Wason y Evans, 1975; Evans y Wason, 1976; Nisbett y Wilson, 1977; Ericsson y Simon, 1980, 1981, 1984; Evans, 1983; Froufe, 1985; Stinessen, 1985; Engelkamp y Zimmer, 1983; Galotti, Baron y Sabini, 1986), queremos resaltar únicamente, dejando de lado los problemas típicos de la generalización inductiva, la dificultad que entraña la utilización de un instrumento metodológico cuya utilidad viene avalada precisamente por un enfoque teórico que busca contrastación empírica en los propios protocolos verbales y que, además, determina la forma de análisis de los datos obtenidos. De alguna manera, esto equivale a tomar como punto de partida ciertos presupuestos teóricos, los cuales son también los que determinan qué datos pueden ser considerados relevantes y válidos para la contrastación del propio esquema inicial,

lo cual implica aceptar un criterio de validez teórico y no metodológico.

No terminan aquí los problemas para el establecimiento de una equivalencia funcional fuerte entre dos sistemas. Como se ha señalado repetidamente, es absolutamente necesario establecer un conjunto de operaciones primitivas, ejecutadas por la arquitectura funcional del sistema, que nos permitan distinguir entre ejecuciones realmente equivalentes y simples simulaciones. En caso contrario, nos arriesgamos a confundir lo que, en los términos utilizados por Rosen (1986), podemos denominar "modelización" (que preservaría la estructura) y "simulación" (que no preservaría la estructura). En un nivel superior, es también importante distinguir entre procesos de computación distintos en términos de sus objetivos (probar una hipótesis, recodificar una cadena de símbolos, etc.) o de algunas de sus propiedades, a fin de alcanzar el viejo objetivo de individualizar los procesos psicológicos. Como señala Maloney (1985), este es uno de los grandes fracasos de la psicología cognitiva, tal y como está planteada. En cualquier caso, a falta de esta individualización, la posibilidad de contrastación de una equivalencia funcional fuerte se convierte en algo harto dudoso.

Como se ha señalado en repetidas ocasiones, la visión funcionalista propia de los modelos computacionales parte de la base de que el substrato material en el cual se implante cierto sistema computacional es relativamente irrelevante



(Fodor, 1981b); de esta forma, el nivel de implementación material propuesto en la sistematización de Marr quedaría fuera del interés de los psicólogos, salvando el aspecto metodológico de la búsqueda de indicadores (Arnau, Honrubia y Viader, 1985). Siguiendo a Johnson-Laird (1983), el único requisito que cabe exigir al substrato material es su capacidad para ejecutar funciones recursivas puesto que, de acuerdo con la tesis de Church, la recursividad es la característica fundamental de las funciones computables. A nuestro entender, esta posición es difícilmente discutible si nos mantenemos al nivel de teoría de la computación (es decir, a nivel de competencia); el problema es que, cuando intentamos hallar algún tipo de criterio empírico que nos permita distinguir entre diversos modelos computacionales nos vemos obligados a hacer referencia a propiedades de la arquitectura funcional del sistema (en particular, a la definición de sus operaciones primitivas). Dando un paso más, nos parece difícil poner en duda que esas propiedades de la arquitectura funcional se relacionan bastante directamente con la **arquitectura material** del dispositivo de computación utilizado. Tomemos de nuevo el caso de la máquina de Turing. Como se indicaba algunas líneas más atrás, una máquina de este tipo presenta una relación cuadrática entre el número de símbolos que debe leer y el tiempo de lectura. Parece fácil imaginar un dispositivo computacional que, utilizando incluso los mismos elementos básicos, pueda mostrar una relación distinta. Por ejemplo, el cabezal de lectura podría

no ser único, sino estar compuesto por varios elementos capaces de "leer en paralelo" y conectados entre si. No nos parece necesario entrar en detalles técnicos para ver claramente que la relación establecida entre número de símbolos y tiempo de lectura será distinta a la encontrada en la máquina original; si la relación encontrada fuera lineal (o si, como en las máquinas de registro, no existe tal relación), podremos hablar de diferencias "cualitativas" entre los dispositivos computacionales, y no solamente de diferencias corregibles por un criterio proporcional, como propone Simon (1978). Aunque esto pueda ser irrelevante para una teoría de la competencia (para la cual, por otro lado, la mayoría de los criterios de contrastación empírica son poco o nada significativos), esta observación nos parece importante cuando nos planteamos la realidad psicológica de los modelos propuestos y su adecuación a los datos experimentales.

La necesidad de retornar a criterios menos formalistas en la modelización (o de buscar formalismos más fructíferos) se refleja también, por ejemplo, en el nuevo impulso que los criterios de plausibilidad neurológica están adquiriendo en el marco de la psicología cognitiva. Aunque podrían citarse diversos ejemplos de esta tendencia, nos parece especialmente relevante la recopilación de trabajos realizada por Hinton y Anderson (1981), en la cual la posibilidad de "implementación" de los modelos propuestos en el "hardware" humano juega un papel tan importante como las ideas de procesamiento paralelo o estructuras distribucionales.

Resumiendo la idea que estamos intentando exponer, puede decirse que el funcionalismo, en tanto que doctrina filosófica, sugiere que es posible (pero, añadimos nosotros, ni necesario ni probablemente conveniente) ignorar, a ciertos niveles, el aspecto de implementación material de un modelo computacional. En nuestra opinión, esto es perfectamente coherente siempre que nos movamos en el nivel de teoría de la computación, pero se hace mucho más dudoso en el momento en que descendemos a un nivel algorítmico el cual, como estamos viendo y como se desprende también del esquema original de Marr, es el nivel propiamente psicológico.

Consideremos la situación que se nos plantea cuando podemos trabajar con una cierta variedad de computadores. Algunos de ellos pueden no poseer la distinción entre procesador central y memoria, otros pueden trabajar en paralelo y no en serie, etc. Si construimos un programa para la ejecución de cierta tarea, lo primero que debemos plantearnos es la clase de máquina para la cual preparamos el programa; de hecho, un programa no es otra cosa que la descripción de la secuencia de actividades que cierto dispositivo debe realizar, y se relaciona directamente con las propiedades funcionales de la máquina las cuales, a su vez, surgen directamente de las propiedades físicas del sistema de computación. No se trata únicamente de diferencias de capacidad o eficiencia. El problema se revela como mucho más fundamental cuando nos planteamos la posibilidad de que la arquitectura del cerebro humano pueda ser distinta de la arquitectura básica

propuesta por von Neumann. Kosslyn y Hatfield (1984) señalan algunas de las diferencias fundamentales: distinción CPU/memoria versus módulos relativamente independientes; diferencias en el formato de instrucciones y datos; almacenamiento "espacial" frente a almacenamiento "difuso" o distribucional; diferentes niveles de "programabilidad"; procesamiento serial versus paralelo. Al margen de que todas estas cuestiones estén a debate en el ámbito de la fisiología (aunque parece haber un cierto consenso en cuanto a algunas de ellas), es indudable que este tipo de restricciones introduce consecuencias empíricas que no pueden ser ignoradas en el trabajo de investigación realizado en este campo.

## 2.2 Estructura y naturaleza de los símbolos

Los modelos computacionales que venimos describiendo presuponen, como se ha indicado repetidamente, la existencia de unidades atómicas que constituyen la materia prima de las operaciones computacionales. Estas unidades últimas son concebidas como símbolos atómicos, discretos, no estructurados, comparables y copiables como, por ejemplo, los dígitos que imprime una máquina de Turing en su correspondiente cinta de papel. La comparación entre dos símbolos atómicos permite definirlos como iguales a distintos (sin gradación posible, dado su carácter inestructurado), lo cual hace posible la identificación de clases de símbolos y la introducción del concepto de símbolo-tipo. Además, mediante la introducción

de ciertas relaciones simples (por ejemplo, la relación "siguiente") es posible construir secuencias de símbolos que constituyen unidades complejas para el procesamiento.

Debe señalarse, no obstante, que en muchas aplicaciones concretas del enfoque computacional se tiende a definir un alfabeto propio de símbolos "primitivos", aun a sabiendas de que no se trata de los símbolos más elementales (Newell y Simon, 1972). Por ejemplo, si se trabaja en solución de problemas puede ignorarse el trabajo de procesamiento necesario para el reconocimiento de letras y tomar éstas, junto con los números y una serie de símbolos especiales tomados de la lógica, como símbolos primitivos, aun reconociendo la posibilidad de una descomposición de tales símbolos en constituyentes atómicos. En términos generales, pueden tomarse como "primitivos" aquellos símbolos que constituyan la materia prima de las operaciones elementales de procesamiento relevantes para el nivel de análisis que se esté realizando; desde el momento en que cualquiera de estas estructuras simbólicas puede ser descompuesta en elementos atómicos podremos definir las como **analíticas**, y a partir del carácter no estructurado y discreto de esos elementos atómicos, como **copiables o reproductibles**. Otra cuestión distinta es que la falta de definición de un conjunto de operaciones elementales de procesamiento dificulte la identificación de las estructuras simbólicas básicas para cualquier nivel de análisis de los procesos cognitivos.



Naturalmente, la clase de símbolos propia de los modelos computacionales no agota el conjunto de todos los tipos de símbolos posibles. Sin pretender analizar las teorías de la simbolización y, en particular, sin introducir referencias profundas a la semiótica (o semiología, para hacer justicia tanto a Peirce como a Saussure), si nos parece necesaria la revisión de algunos trabajos relevantes al tema que estamos tratando. En particular, la teoría de la notación de Goodman (1968) ha sido propuesta como base para un modelo de simplicidad estructural (Krueger y Osherson, 1980), y también como punto de partida tanto para la construcción de modelos computacionales (Pylyshyn, 1973, 1986) como para su crítica (Kolers y Smythe, 1979, 1984). Este tipo de crítica parte de la base de que los símbolos utilizados por los teóricos computacionales no son sino un subconjunto bastante restringido de todas las clases posibles de símbolos, y en ella se analiza la posibilidad de que el sistema cognitivo humano funcione en base a tipos de símbolos distintos a los propios de las teorías computacionales. Analizaremos este último aspecto con cierto detalle.

La idea básica que puede extraerse de la teoría de Goodman es la distinción entre símbolos articulados y símbolos densos, es decir, entre símbolos que son perfectamente copiables y aquellos que no poseen esa cualidad. Se trata de una distinción semejante a la que puede establecerse, en el campo de las artes, entre artes copiables (alográficas) y no copiables (autográficas). Un tema musical, por ejemplo,

puede considerarse generalmente como un ejemplo de producto alográfico, puesto que puede ser reproducido perfectamente a partir de su partitura; en este sentido, serán irrelevantes todo un conjunto de propiedades tales como el color de la tinta en que se reproduzca esa partitura, el tipo de grabación realizado (disco, cinta de cassette...), etc. En cambio, una pintura constituye un ejemplo de producto autográfico y denso, es decir, una unidad simbólica que no admite modificaciones en ninguno de sus aspectos. En otros términos, una partitura (o un poema) son copiables desde el momento en que poseen una notación compuesta de símbolos primitivos y discretos; una pintura no es copiable, puesto que no disponemos de tal notación. Naturalmente, es fácil darse cuenta de que los modelos computacionales toman como elementos básicos lo que Goodman llamaría símbolos alográficos, mientras que otros modelos parten de la consideración de estructuras simbólicas densas (por ejemplo, los modelos propuestos por algunos teóricos de la imagen mental).

En tanto que teoría de la referencia, el trabajo de Goodman se centra en una clasificación de los símbolos en cuanto a la fiabilidad de su función referencial; de acuerdo con su análisis, esa fiabilidad es máxima cuando los símbolos poseen ciertos criterios sintácticos en cuanto a su estructura y ciertos criterios semánticos en cuanto a la forma en que representan el entorno. En ambos casos el aspecto principal es el grado de diferenciabilidad, lo cual nos conduce

inmediatamente a la cuestión de la densidad de la estructura simbólica, tanto en su aspecto sintáctico como semántico. La densidad sintáctica se refiere a la naturaleza del propio símbolo (por tanto, si aceptamos el punto de vista de la psicología cognitiva, a la naturaleza de la representación en tanto que sistema simbólico); la densidad semántica se refiere a la naturaleza del dominio que esté siendo representado. En principio, no parecen existir grandes dificultades en cuanto a la densidad semántica: tanto una representación densa como otra de tipo discreto son capaces de representar aspectos de un dominio semánticamente denso; por ejemplo, un círculo puede ser representado mediante una imagen o a través de la correspondiente palabra aunque, como señala Dretske (1981), en el caso de la representación densa poseemos siempre una cantidad de información suplementaria potencialmente irrelevante (que proviene, en realidad, del carácter irreductiblemente particular y concreto de las imágenes). Pero, a pesar de la aparente claridad de esta cuestión, creemos necesario introducir una precisión que nos parece fundamental. Tomemos como ejemplo el de la contraposición, frecuentemente establecida, entre lenguaje natural y lenguaje matemático (Thom, 1980). Consideremos un oscilador lineal, y construyamos una descripción lingüística de su comportamiento. Obviamente, el lenguaje matemático nos permitirá una descripción mucho más precisa, basada quizás en la introducción de un espacio de fases y en el establecimiento de un conjunto de ecuaciones correspondientes a

trayectorias elípticas en torno al origen. En este sentido, puede afirmarse con rotundidad que la descripción lingüística no es sino una aproximación bastante imperfecta al fenómeno que estamos analizando y, en particular, resulta claro que no puede aprehender ciertas propiedades relacionadas con el carácter intrínsecamente continuo de la dinámica del oscilador. Pensamos que esta idea tiene bastante interés, puesto que, como veremos en el capítulo 4, nos presenta una situación comparable en cierto sentido con la que se plantea al intentar relacionar un substrato dinámico de carácter continuo (al menos parcialmente) con un sistema notacional basado en elementos discretos. Debe señalarse, además, que las dificultades de los códigos discretos ante los fenómenos de carácter continuo constituyen el punto de partida del proyecto de "geometrización del significado" propuesto por Thom (1972, 1982).

Naturalmente, cualquier teórico computacional podría responder a estas argumentaciones aportando el caso del cálculo diferencial, en tanto que ejemplo representativo del análisis de procesos continuos a través de operaciones computacionales discretas. Esto es cierto hasta el punto de que Newell y Simon (1972) sugieren que el programa de computador juega, en el marco de sus modelos, un papel semejante al del cálculo diferencial: se trata de procedimientos de "discretización" de la realidad, mediante los cuales lo que originariamente podía ser continuo resulta abordable a través de un conjunto de pasos discretos. Algo parecido podría decirse

del tránsito, comentado en el capítulo 1, de la "geometría densa" clásica a la geometría analítica. Permítasenos, sin embargo, decir algo que nos parece importante: lo que demuestra este tipo de ejemplos es una peligrosa tendencia a confundir las propiedades de los objetos de estudio con las características de los instrumentos conceptuales destinados a su análisis. Aunque un computador digital puede, mediante el cálculo diferencial, determinar la posición de un móvil en un momento  $t$  (y, además, con una precisión arbitraria, aunque naturalmente no infinita), suponemos que ningún físico se atreverá a proponer una "equivalencia funcional fuerte" entre el comportamiento del ordenador y el del móvil objeto de estudio. Estamos intentando sostener una argumentación semejante para el caso de los procesos cognitivos humanos; como señala Rivière (1986) no pueden confundirse las representaciones del psicólogo con las del sujeto, ni atribuir a éste competencias recogidas en modelos para los cuales no resulta posible, como argumentábamos en el apartado anterior, acceder al requisito de equivalencia fuerte. El caso de la densidad sintáctica resulta también problemático, aunque en un sentido algo distinto. Como señala Pylyshyn (1983), la utilización de símbolos analíticos (dotados, por tanto, de diferenciación o articulación sintáctica) se encuentra en la base de todos los sistemas notacionales, y es fundamental para la matemática en general, y en particular para toda la formulación surgida del trabajo de Turing. Hablar de un sistema sintácticamente denso exige referirse a



un sistema simbólico en el cual no existen símbolos atómicos, de forma que cualquier par de símbolos puede diferir en un grado arbitrariamente pequeño. De esta forma, puede afirmarse que toda diferencia entre objetos semánticos podría estar descrita en los objetos sintácticos; la falta de discriminación entre dos objetos distintos podría producirse únicamente a consecuencia del limitado poder discriminativo de nuestros órganos sensoriales. En este sentido, y como reconoce el propio Pylyshyn (1983), no pueden rechazarse las representaciones densas porque contengan una cantidad excesiva (en realidad, potencialmente infinita) de información. En el cerebro existe un número enorme de procesos físico-químicos continuos que pueden enfrentarse con representaciones densas. En realidad, lo fundamental es la "interpretación" de la representación. De acuerdo con la teoría de la información, la cantidad de información (y propiedades ligadas a ella, como la diferenciabilidad), no es algo inherente al objeto, sino que depende de cómo se "interprete" éste, es decir, entre cuantos objetos distintos pueda distinguir el intérprete. El requerimiento de finitud en la cantidad de información indica sencillamente que el intérprete no puede distinguir entre una cantidad indefinida de configuraciones, pero no implica que estas configuraciones tengan que ser forzosamente discretas.

En esta misma línea de "relativización", debe señalarse que la característica de copiabilidad no es absoluta: por ejemplo, un dibujo sencillo puede ser descrito lingüísticamente

de forma exhaustiva y, en consecuencia, resulta ser reproducible. Sin embargo, cuando la complejidad aumenta esta traducción se hace inviable, y sólo es posible reproducir aspectos parciales de la estructura simbólica densa. Lo que sí es posible es utilizar otros medios digitales, por ejemplo, algún tipo de "filtro" que active una "alarma" cuando observe una situación absolutamente idéntica a la expresada en un dibujo. Este tipo de mecanismo es importante en el reconocimiento de formas por ordenador. Naturalmente, no será posible llegar a un grado de precisión infinito pero, como se acaba de indicar, tampoco nuestros órganos sensoriales (ni, presumiblemente, nuestro "ojo de la mente") poseen tal nivel de exactitud. Debe señalarse, en todo caso, que cualquier transformación desde un dominio denso a un sistema notacional discreto (en caso de que sea posible) implicará la pérdida del carácter análogo de la representación densa. En este sentido puede afirmarse que la condición de "copiabilidad" es realmente fuerte y, de acuerdo con Kolers y Smythe (1984), justifica la apelación al carácter personal o autográfico de las representaciones densas.

Al margen de que la posición de Kolers y Smythe posea importantes puntos débiles (que han sido analizados, por ejemplo, por Allport, 1984, o por Brooks, 1984), la insistencia en el recurso a representaciones densas encuentra su justificación en las insuficiencias de los sistemas simbólicos discretos en tanto que modelos de la representación. Aunque nos hemos referido ya a importantes

problemas de los modelos computacionales (especialmente en cuanto a su carácter de teoría de la competencia) , pensamos que existen otras cuestiones que atañen más directamente a las dificultades que un código representacional discreto encuentra ante determinados aspectos que debe aprehender necesariamente una teoría psicológica fuerte (y que, eventualmente, podrían hallar sentido en el marco de modelos alternativos de la representación). Pensamos igualmente que muchos de estos aspectos se refieren a lo que, en términos generales, podríamos denominar "flexibilidad conductual", es decir, a la capacidad del sistema cognitivo humano para adaptarse a nuevas situaciones, realizar cambios en su propia estructura como resultado del aprendizaje, solucionar problemas, etc. Dada la relevancia de estas cuestiones, trataremos algunas de ellas en el apartado 2.3. Adelantaremos, no obstante, la tesis de que los modelos computacionales discretos deben incorporar a-priori prácticamente todas las posibilidades de complicación del entorno, lo cual les confiere un carácter profundamente innatista y considerablemente estático.

Ocupémonos ahora de otras cuestiones. Como es sabido, la oposición de los teóricos proposicionales a las representaciones análogas y, en particular, al valor funcional de las imágenes mentales, radica básicamente en razones de tipo lógico. En particular, Pylyshyn (1981) razona su afirmación sobre el carácter epifenoménico de las representaciones en forma de imágenes a partir de su conocido argumento sobre

los niveles explicativos. Como se indicaba ya en el capítulo anterior (aunque en términos algo distintos), Pylyshyn postula la existencia de dos tipos básicos de explicación para la conducta de un sistema: el primer tipo de explicación apelaría a propiedades intrínsecas del sistema; sería suficiente, en este caso, con explicitar el conjunto de leyes naturales que rigen esas propiedades o, en casos más complejos, con desarrollar explicaciones semejantes a las de la ingeniería en cuanto a la forma en que la interacción de diversos componentes del sistema da lugar a las conductas observadas (Haugeland, 1978). El segundo tipo de explicación -que sería, de acuerdo con los teóricos computacionalistas, el verdaderamente relevante para la psicología cognitiva- acudiría a propiedades "extrínsecas" al propio medio de computación y, más concretamente, a propiedades y conductas de los "mundos" con los cuales el sistema mantiene cierta relación (relación que, de acuerdo con los teóricos computacionales, es de tipo semántico pero que, como hemos indicado ya en el capítulo 1 y como argumentaremos en el apartado 2.4, no es sino una relación sintáctica más). El tipo de explicación más adecuado dependerá esencialmente de la naturaleza de las regularidades conductuales observadas. Como se indicaba ya en el capítulo 1, Pylyshyn sostiene que solamente el segundo tipo de explicación es adecuado para aprehender las regularidades observadas en el funcionamiento del sistema cognitivo (en particular, para dar cuenta de las regularidades que son independientes del formato físico preciso que

tome la representación); de acuerdo con este punto de vista, el primer nivel explicativo, al cual pertenecerían los modelos de imágenes mentales, sería insuficiente desde el momento en que las propiedades del medio de computación influirían decisivamente en la conducta del sistema, de forma que sería imposible aprehender lo que de funcionalmente equivalente puedan tener las diversas concreciones físicas de una misma representación.

La necesidad de recurrir al nivel de las representaciones semánticamente interpretadas se pone de manifiesto, de acuerdo con Pylyshyn (1973), en el carácter cognitivamente penetrable de las imágenes mentales. El hecho de que la manipulación de imágenes se vea influida por las creencias, metas y conocimiento tácito de los sujetos demostraría la existencia de una base representacional común que, como se ha indicado repetidamente, poseería un carácter analítico y sería inaccesible a la conciencia. Sin embargo, debe señalarse que el criterio de penetrabilidad cognitiva no está exento de problemas. Por ejemplo, resulta imprescindible introducir algunas restricciones en su misma definición, puesto que en caso contrario nos vemos obligados a tratar como cognitivos una serie de fenómenos que parecen tener poca relación con nuestro ámbito de estudio. Por ejemplo, la afluencia de sangre al cerebro es afectada por las creencias del sujeto en cuanto a los estímulos que está percibiendo; ¿debemos pensar entonces que el funcionamiento del sistema circulatorio está regido por reglas expresadas



proposicionalmente?. Y si la respuesta es negativa, ¿cual es el criterio que permite diferenciar entre un proceso descrito por reglas (como puede serlo, obviamente, el funcionamiento del sistema vascular) y un proceso dirigido por reglas?. ¿Existe realmente este último tipo de proceso?.

A pesar de las dificultades del enfoque computacional, la respuesta de los teóricos de la imagen a algunas de sus argumentaciones no ha sido habitualmente satisfactoria, al menos a nivel lógico (aunque, a nivel empírico, existe un corpus de evidencia lo bastante amplio como para apoyar fuertemente su posición). Tanto es así, que algunos autores muy significativos (en especial, S.M. Kosslyn) parecen haber oscilado entre una posición "conciliadora", en la cual el criterio básico de apoyo para las imágenes no era ya su relativa "continuidad" respecto a los procesos perceptivos sino la posibilidad de construir algoritmos de computación que simulen sus propiedades (Kosslyn y Schwartz, 1977; Kosslyn, 1980, 1981), y una posición mucho más radical en la que se llega a negar la existencia de sistemas simbólicos subyacentes a la representación (Kosslyn y Hatfield, 1984).

En particular, en la primera de estas alternativas la generación de representaciones superficiales transitorias (imágenes) se produce en un "buffer" visual en base a la información "profunda" almacenada en la memoria a largo plazo. Esta información puede estar codificada de forma "literal" o de forma proposicional (Kosslyn, 1981). La codificación literal conservaría la estructura básica del estímulo (en

particular, su forma global y la situación de regiones locales de los objetos), de forma que una imagen podría ser generada por activación de su codificación literal correspondiente. Las diversas partes ("regiones") de un objeto codificado literalmente (y también los distintos objetos) están unidas por relaciones proposicionales que especifican a qué objeto pertenecen cada una de las regiones y las relaciones entre los diversos objetos. De acuerdo con Kosslyn (1980, 1981), estas relaciones son necesarias a fin de explicar la flexibilidad con que las imágenes pueden ser reorganizadas y combinadas entre sí.

Es particularmente ilustrativo observar las dificultades de la posición de Kosslyn en cuanto a mantener la relevancia de las propiedades analógicas de la representación. Puede señalarse, por ejemplo, que la codificación "literal" de la información se realiza mediante unidades no espaciales; en realidad, la información toma la forma de listas de coordenadas que especifican la situación que los diversos componentes del objeto codificado deben tener en una "matriz superficial" que describe la imagen de ese objeto. Aunque Kosslyn insiste en que estos detalles de implementación no son relevantes para el esquema teórico global, debemos señalar que nos encontramos ante una de las dificultades básicas de los modelos que recurren a la simulación en ordenador: la problemática distinción entre presuposiciones teóricas y aspectos de programación y codificación de información en el ordenador. Por otra parte, si este tipo de codificación

digital es capaz de almacenar la información relacionada con las imágenes, ¿por qué (y cómo) insistir en la relevancia de las propiedades del medio representacional?

Una idea muy parecida es expuesta por Yuille (1983) cuando se refiere a la auto-contradicción que parece implicar la postura que estamos exponiendo. Específicamente, Kosslyn (1980) presenta un modelo computacional de la imagen en el cual la información analógica que supuestamente posee una imagen es desplegada en una pantalla. Posteriormente, este despliegue sirve de base para la simulación de los procesos que hipotéticamente subyacen a la comparación de tamaños, juicios de distancia, rotación mental, etc. El problema es que el aspecto analógico de lo representado en pantalla no juega ningún papel en la realización de las anteriores tareas por el computador. La información que el computador utiliza no es analógica; sino que ha sido traducida anteriormente a información de tipo digital, de forma que las características analógicas de la imagen no producen ningún efecto, como tales, en la ejecución. Parece probable que la postura adoptada por Kosslyn en algunos trabajos posteriores (en especial, en el artículo de Kosslyn y Hatfield, 1984) no sea ajena a algunas de estas consideraciones.

En relativo contraste con los modelos "mixtos" se encuentran las teorías "puras" de la imagen, de entre las cuales puede destacar por su nivel de desarrollo la teoría de código dual de Paivio (Paivio, 1971, 1979; Paivio y Begg, 1981; Yuille, 1983). Dado su carácter casi clásico, obviaremos una

exposición detalla de la teoría y nos centraremos en los aspectos que puedan ser más relevantes para la problemática que estamos considerando. (Ver figura 2.2).

Como es bien sabido, la teoría de código dual rechaza la existencia de cualquier forma de representación de tipo amodal (y, en particular, del formato proposicional) y postula, en cambio, la presencia de dos sistemas de codificación independientes, aunque estrechamente relacionados: el sistema de imágenes, en el cual se incluirían las representaciones densas que estamos considerando hasta el momento, y un sistema verbal que comprendería verbalizaciones internas que, en última instancia, consistirían en "imágenes" auditivas o fono-articulatorias.

Debe señalarse, en primer lugar, que los modelos basados en imágenes mentales han recibido un nivel de apoyo empírico bastante considerable. Un amplio conjunto de experimentos ha aportado información sobre las propiedades de este tipo de representaciones y, como veremos inmediatamente, algunos de los datos obtenidos, aun siendo interpretables desde un modelo computacional discreto (en rigor, cualquier dato es asimilable por este esquema), obligan a este tipo de modelos a realizar presuposiciones progresivamente menos verosímiles. Citaremos esquemáticamente algunos datos experimentales de interés: 1) Los juicios de similaridad entre objetos son básicamente iguales si los objetos están presentes efectivamente o si únicamente son nombrados (Shepard y Chapman, 1970; Shepard y Castro, 1975).

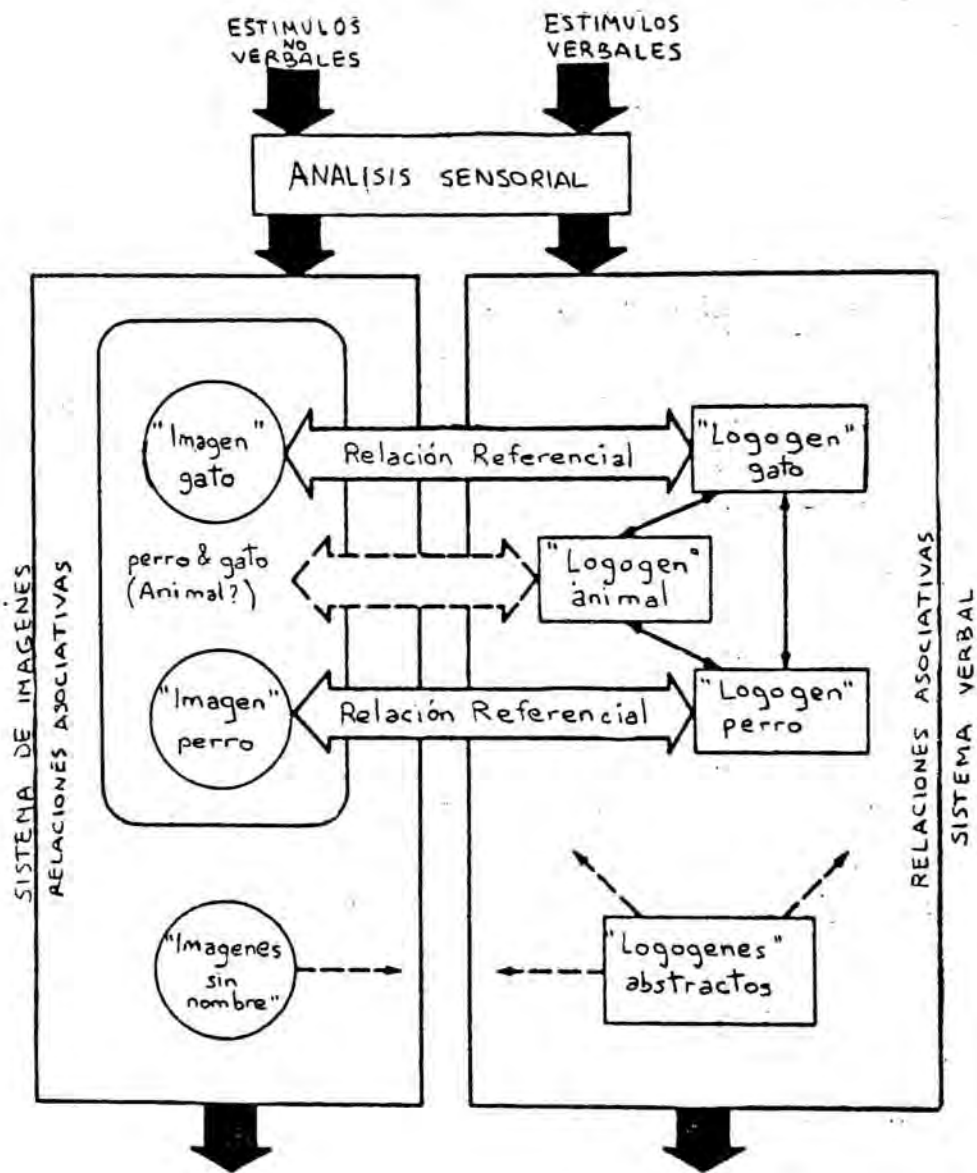


Figura 2.2 Esquema general de la teoría de código dual



- 2) La "forma" de los nombres o símbolos utilizados en lugar de los objetos mismos no tiene efecto sobre los juicios de semejanza (Shepard, Kilpatrick y Cunningham, 1975)
- 3) El tiempo necesario para determinar que un objeto es del mismo tipo que otro objeto estándar aumenta monotónicamente a medida que crece la diferencia transformacional entre ambos objetos (Cooper y Shepard, 1975; Cooper, 1975; Shepard y Feng, 1972)
- 4) Existe conservación del "momento representacional", pudiendo hablarse, de alguna forma, de "inercia" de las imágenes mentales (Freyd y Finbke, 1984)
- 5) En una tarea de elección forzada, la inducción de imágenes "apareadas" con el estímulo externo produce una mejor detección. Una imagen apareada facilita la detección solamente cuando imagen y estímulo están en la misma posición (Farah, 1985)
- 6) La imaginación de un movimiento puede interferir la ejecución de tareas motoras (Johnson, 1982)
- 7) El tiempo de reacción para determinar si un número es mayor o menor que otro ya presentado disminuye monotónicamente a medida que crece el tamaño de la diferencia entre ambos (Schuler, Rubin y Armstrong, 1971; Moyer y Landauer, 1973)
- 8) La información estructural contenida en una representación interna es accesible al sujeto para futuras operaciones en ausencia del correspondiente estímulo externo (Cooper, 1973)

Shepard (1975, 1978, 1984) sugiere que la evidencia disponible hasta ese momento juega en favor de un sistema representacional basado en imágenes mentales; en este contexto, la representación interna no consiste (o no consiste solamente) en una pronunciación sub-vocal o "mental" del nombre del objeto, ni una descripción verbal implícita del mismo (Shepard y Castro, 1975; Metzler y Shepard, 1974), ni una "impresión" de ciertas características detectadas en el input sensorial, ni un modelo concreto, espacialmente isomorfo. Este último punto es particularmente importante, puesto que introduce la noción de isomorfismo de segundo orden, básica para los teóricos de la imagen.

En principio, la noción de isomorfismo de segundo orden no es excesivamente precisa, y puede ser definida más en términos negativos (en contraposición con el isomorfismo de primer orden) que a partir de conceptos "positivos". Se trata, esencialmente, de un isomorfismo funcional, y no de un isomorfismo físico; en otras palabras, cuando una persona dice que "esto es una elipse", en su cerebro suceden hechos semejantes a los que se producen en otro cerebro ante la misma circunstancia, sin que esto implique semejanza estructural entre el objeto externo y los procesos internos. De alguna forma, la similitud entre los objetos externos debe reflejarse, en un isomorfismo de segundo orden, en la similitud entre los objetos correspondientes que los representan, pero no en una semejanza entre objetos externos y los procesos que dan lugar a la representación. El gran problema se

plantea en el momento de definir en qué consiste este "isomorfismo abstracto" entre sucesos cerebrales y objetos externos; en este sentido, los teóricos de la imagen han hecho poco más que recurrir a metáforas más o menos difusas como la de la llave y la cerradura (Shepard, 1978). Posteriormente introduciremos algunas ideas que nos parecen interesantes a este respecto. Básicamente, como se verá en los capítulos 4 y 5, se tratará de acudir a un criterio de isomorfismo basado en consideraciones topológicas; de acuerdo con este criterio, la equivalencia entre dos representaciones (o entre un objeto y su representación) se establecería en función de la semejanza entre lo que denominaremos sus "figuras de regulación".

La teoría de código dual que, como hemos visto, puede ser considerada también como teoría "pura" de la imagen, ha generado su propio corpus de datos experimentales, que en buena parte refuerzan los resultados obtenidos en el marco de los experimentos tradicionales de manipulación de imágenes mentales. Como señala Yuille (1983), el conjunto de datos experimentales recogidos por los autores que trabajan en esta línea es extraordinariamente extenso y a veces técnicamente complejo, por lo que nos limitaremos a señalar algunos de los hallazgos más relevantes. Estos pueden referirse, esencialmente, a los efectos facilitadores de la concreción del material experimental (interpretados en términos de la mayor o menor facilidad para la formación de imágenes a partir de distintas palabras), efectos estadísticamente

independientes de imágenes y palabras en recuerdo libre, efectos de las instrucciones de codificación, efecto de distancia simbólica, interferencia específica para cada una de las dos modalidades, etc. (Paivio y Csapo, 1969, 1973; Paivio, 1971, 1978, 1983; Yuille, 1983; Moyer y Bayer, 1976, etc.). Resulta particularmente ilustrativa la recopilación de literatura experimental realizada por Paivio (1983), quien cita un total de 60 hallazgos experimentales bien establecidos que juegan en favor de la teoría de código dual. De entre el conjunto de datos aportados por Paivio (1983) quisiéramos destacar un hecho al que atribuimos una importancia crucial. Se trata del efecto de distancia simbólica (Moyer y Bayer, 1976) que, de hecho, puede ser interpretado como una generalización de alguno de los resultados a los que nos referíamos anteriormente. De acuerdo con Moyer y Bayer, el tiempo de reacción necesario para realizar un juicio comparativo de dos objetos respecto de cierto atributo disminuye monotónicamente a medida que la diferencia en ese atributo aumenta. Este hallazgo juega muy poderosamente en favor de la existencia de representaciones que incorporan características de tipo continuo.

La dificultad básica de las teorías de la imagen reside, a nuestro entender, en el problema de la generalización y la abstracción. En la línea de Kolers y Smythe (1979, 1984), debemos resaltar el carácter autográfico y episódico de las imágenes mentales, lo cual plantea inmediatamente la cuestión de la emergencia de conceptos de orden jerárquico

superior. Si se acepta, de alguna forma, que las actividades conceptuales tienen su origen último en la percepción, debemos preguntarnos cómo de perceptos concretos pueden derivarse conceptos abstractos. En esta línea, Jusczyk y Earhard (1980) responden negativamente a la cuestión que ellos mismos plantean: ¿Puede un modelo de código dual no ser proposicional?, es decir, ¿es posible llegar a la abstracción sin recurrir a un sistema representacional amodal?. Después de introducir interpretaciones proposicionalistas de los datos aportados por la teoría de código dual (algo que, en realidad, no resulta demasiado difícil, dada la versatilidad del formato proposicional), Jusczyk y Earhard recogen el argumento clásico de los teóricos computacionales al abogar en favor de la existencia de un "lenguaje de la mente" de carácter amodal que permita "interpretar" las cadenas verbales haciendo abstracción de sus propiedades particulares (aspectos fonéticos, peculiaridades gramaticales, etc.). Sin embargo, pensamos que es posible propugnar un mecanismo abstractor sin necesidad de postular un lenguaje de la mente de formato proposicional. Este sistema abstractivo se basaría en procesos que presentan un carácter parcialmente continuo, de forma que, aunque admitiera (como cualquier otro proceso) una descripción proposicional en términos de sentencias y reglas, ésta no sería sino una aproximación imprecisa que, además, no podría cumplir de ninguna forma con el criterio de "equivalencia funcional fuerte" que analizábamos en el apartado anterior. Dado que en el capítulo 4



estudiaremos con detenimiento un conjunto de modelos que, entre otras características, poseen una considerable capacidad para incorporar procesos abstractivos, nos limitaremos a señalar aquí que el común denominador de tales modelos es el de postular algunos mecanismos tendentes a potenciar los aspectos comunes a diversos perceptos y a "compensar" las características particulares. En general, esto puede lograrse a través de cambios en patrones de activación continuos asociados a los distintos perceptos. Debe señalarse, además, que este tipo de modelos permitirá retener también las características particulares de los items, sin que por ello interfieran en la información de carácter más general.

Una idea subyacente al tipo de planteamiento que defendemos en cuanto a los procesos abstractivos es que, de alguna forma, la abstracción es un problema de grados. Wickelgren (1972) propone que un concepto sea considerado como una disyunción de conjunciones. De acuerdo con este planteamiento, el problema del grado de abstracción de los conceptos sería el del grado de inclusividad de la regla disyuntiva. Las representaciones más abstractas incluyen disyunciones altamente inclusivas (por ejemplo, imágenes y palabras referidas a un mismo objeto), mientras que las representaciones más "superficiales" serían más restrictivas (admitirían, por ejemplo, solamente imágenes visuales o "imágenes" auditivas). De esta forma, un sistema complejo de conexiones cada más selectivas podría ser una buena aproximación al problema de la abstracción. En torno a esta idea de potenciación de

elementos comunes converge todo un conjunto de trabajos que van desde los realizados por Hebb desde consideraciones fisiológicas (Hebb, 1980) hasta los denominados "modelos neoconexionistas" (Hinton y Anderson, 1981). Como veremos en el capítulo 4, se han obtenido ya algunos resultados estimables en esta línea, lo cual nos hace pensar que nos hallamos ante un planteamiento básicamente correcto.

### 2.3 El "cambio cognitivo" desde una perspectiva computacional

Intentaremos demostrar, en este apartado, que el carácter analítico-discreto del sistema simbólico que propugnan los teóricos computacionales introduce restricciones inaceptables en sus propios modelos. Como señalábamos en el apartado anterior, algunas de estas restricciones se refieren a la dificultad que este tipo de modelos encuentra para enfrentarse al problema general del "cambio", ya sea en el aspecto del aprendizaje, de la solución de problemas o, más en general, a la cuestión de las transformaciones del propio sistema cognitivo ante las exigencias del medio.

Como se indicó ya en otro lugar, los modelos computacionales se enfrentan a dificultades importantes en lo que se refiere al aprendizaje de la lengua. Como señalábamos entonces, el propio Fodor admite la necesidad de la existencia previa de un lenguaje interno que haría posible la asignación de significado a los predicados del nuevo lenguaje externo, y

concluye que solamente el carácter no aprendido de este lenguaje interno puede evitar el peligro de una regresión infinita en la cual la asignación de significados a los predicados de un lenguaje se realizaria por referencia a otro lenguaje anterior. Este tipo de dificultades son comunes al análisis computacional de cualquier tipo de aprendizaje. No resulta sorprendente, en consecuencia, que Fodor (1975, 1981) niegue explícitamente la existencia de aprendizaje en el dominio de lo cognitivo, o que muchos modelos computacionales partan de la base de que el aprendizaje puede ser descrito en términos de la construcción de "secuencias bien formadas" (aceptables según los criterios de la supuesta "gramática de la mente") a partir de procedimientos o conceptos ya existentes, de forma que pueda afirmarse que las nuevas construcciones son reductibles a elementos anteriores, o sustituibles por ellos; es decir, en palabras de Pylyshyn (1978), son "estrictamente hablando, eliminables". ¿Cual es el sentido exacto de este tipo de afirmaciones?

El problema es en buena parte semejante al que plantea la gramática generativa chomskiana (lo cual pone de relieve, una vez más, el estatus de las teorías computacionales como teorías de la competencia), y radica en la supuesta insuficiencia de la estimulación externa para producir nuevos estados representacionales. Una afirmación corriente entre los lingüistas que trabajan en la línea generativa es que el niño se enfrenta continuamente a un ambiente lingüístico que le propone enunciados incompletos o imperfectos, de lo cual

puede concluirse fácilmente que la adquisición de la lengua (y, en especial, de sus reglas sintácticas) deben venir regulada por un conocimiento previo de ciertas reglas de gramaticalidad que permitan clasificar un enunciado como aceptable o no. Debe remarcarse el hecho de que lo que el sujeto posee innatamente es, en este contexto, conocimiento; cuando Chomsky afirma que la teoría de la adquisición del lenguaje es la descripción de cómo madura la facultad del lenguaje, lo que quiere decir es que la ontogénesis de las capacidades lingüísticas es el despliegue de las consecuencias deductivas de las "creencias" innatas (expresadas en forma de proposiciones) en interacción con un conjunto de datos perceptivos.

Un caso bastante distinto, pero relevante para nuestro análisis, puede encontrarse en el estudio de la visión desde la óptica de la inteligencia artificial. Como señala Ullman (1979), en un estímulo lumínico no parece existir suficiente información como para que se produzcan los estados perceptuales a los que dicho estímulo da lugar; en otras palabras, la información contenida en el estímulo no parece suficiente como para determinar de forma unívoca la representación visual. Se hace necesario, entonces, postular la existencia de un conjunto de restricciones o presuposiciones sobre el entorno distal que, como demuestra el propio Ullman, permiten el establecimiento de una única aplicación entre estímulo proximal y representación visual. Lo importante aquí es señalar que estas presuposiciones (por ejemplo, que las

escenas visualmente interpretables contienen básicamente sólidos de tres dimensiones, que las superficies son básicamente continuas, etc.) poseen en este caso (y a diferencia del anterior) el carácter de restricciones localizadas en la arquitectura funcional del sistema, puesto que el funcionamiento del sistema visual parece, en esencia, cognitivamente impenetrable. Una distinción semejante es establecida por Fodor (1983) en su intento de construir una nueva psicología de las facultades (una tarea, por cierto, bastante coherente con el carácter innato que atribuye a las capacidades cognitivas básicas): Fodor distingue entre capacidades de base proposicional (que implican conocimiento tácito previo), y mecanismos psicológicos, que estarían localizados en la arquitectura funcional del sistema y poseerían, por tanto, un carácter no cognitivo.

Pylyshyn (1986) plantea algunas ideas semejantes en su breve análisis del problema del aprendizaje. En realidad, Pylyshyn sitúa esta cuestión en el marco más amplio del "cambio cognitivo", y se plantea cuales serían las formas básicas que ese tipo de cambios podría adoptar. De acuerdo con este autor, la cuestión podría formularse en términos parecidos a los que citábamos en el caso de los teóricos generativistas en lingüística: lo fundamental es demostrar que cierto estado cognitivo ha sido alcanzado a causa de que las condiciones ambientales han aportado una cierta cantidad de información respecto a los hechos del mundo, información a partir de la cual el organismo pueda inferir ciertas creencias



(estados representacionales) cuyo contenido esté relacionado con dichas condiciones. Es decir, para que pueda hablarse de aprendizaje las condiciones ambientales deben proporcionar un grado de información suficiente para hacer posible la construcción lógica del estado representacional correspondiente. En la medida en que sea necesario introducir presuposiciones suplementarias (es decir, en la medida en que intervenga el conocimiento anterior del sujeto), el papel del aprendizaje quedará desvalorizado. Parece introducirse aquí un elemento de gradualidad según el cual la construcción de nuevas representaciones necesitaría recurrir en menor grado al conocimiento previo en función del nivel de información proporcionado por el estímulo. En última instancia, si ciertos eventos ambientales producen un estado representacional y se demuestra, como en el caso de los estudios de Ullman, que el estímulo no contenía la información suficiente y que el proceso es cognitivamente impenetrable, deberá concluirse que las "presuposiciones" suplementarias necesarias para la construcción de la representación no forman parte del conocimiento tácito que posee el sujeto, sino de la arquitectura funcional del sistema, lo cual introduce un elemento no-cognitivo en el ámbito del aprendizaje.

Pensamos, en todo caso, que el planteamiento de Pylyshyn es bastante abstracto y un tanto confuso, especialmente en cuanto a la adscripción de fenómenos al ámbito del conocimiento tácito o al de la arquitectura funcional. En este sentido, debemos hacer aquí una breve referencia a una

observación que realiza el propio Pylyshyn, y que resulta reveladora en cuanto a ciertas dificultades de su planteamiento. Afirma Pylyshyn (1986, pág. 209), al referirse a las restricciones intrínsecas y extrínsecas que determinan la aparición de las regularidades, que "es la arquitectura funcional la que introduce restricciones que producen desviaciones respecto a la omnisciencia". Lo cual equivale, evidentemente, a reconocer que los modelos formales propuestos en el nivel de teoría de la computación son incapaces de introducir restricciones a su excesivo "poder aprehensivo". Pensamos que la cita de esta frase nos dispensa de cualquier consideración ulterior sobre esta cuestión.

Los sistemas de producción parecen ofrecer una vía menos abstracta de aproximación a ciertos comportamientos modulados por habilidades. Este tipo de sistemas constituye un ejemplo de lo que Kolers y Smythe (1984) denominan "modelos relacionales", en el sentido de que se expresan en forma de relaciones gobernadas por reglas (esencialmente, de relaciones del tipo si...entonces). Aunque la incorporación directa de principios de aprendizaje en los sistemas de producción se encuentra todavía en sus comienzos, parece existir entre muchos teóricos computacionales la esperanza de que dichos sistemas constituyan la base de una aproximación computacional al aprendizaje (Neves y Anderson, 1981). Sin embargo, pueden citarse algunas limitaciones iniciales que pueden hacer dudar de tal idea (Lewis, 1978). En principio, es perfectamente claro que el programador humano es quien, en el

papel genuino de una especie de "deus ex maquina" introduce nuevas producciones, y los escasos intentos de suprimir la intervención humana parecen muy lejos del éxito. De hecho, un examen cuidadoso de los sistemas de producción propuestos hasta el momento revela su carácter profundamente estático, en la más pura línea del enfoque computacional. Tomemos, por ejemplo, el caso del sistema propuesto por Anderson (1980) para la adición de dos números (tabla 1). Resulta obvio que la serie de producciones que componen el sistema determinan de forma absolutamente mecánica la secuencia de operaciones de procesamiento que deben llevarse a cabo, puesto que a cada condición antecedente se responde con la activación de un sub-objetivo preciso (y del correspondiente algoritmo de ejecución). En la estructura del sistema de producción no se advierten demasiadas posibilidades de complicación sin intervención del programador; podría postularse, por ejemplo, que la condición consecuente de alguna de las reglas de producción fuera la de añadir una nueva regla al sistema. Este es el caso, por ejemplo, del sistema ideado por Waterman (1970), que añade nuevas producciones a medida que se enfrenta a ciertos eventos ambientales específicos. De hecho, este tipo de programas están contruidos en forma de "programas diagnósticos" que detectan situaciones en las cuales algún aspecto del sistema de producción inicial no es satisfactorio, e incluyen un conjunto de reglas de producción destinadas a producir nuevas reglas. En consecuencia, el programa contiene ya inicialmente un conjunto de reglas

Tabla 1. Conjunto de producciones para la suma de dos números. (Anderson, 1976, 1980)

---

- Prod. 1: SI el objetivo es sumar dos números y el primer número termina en un dígito y el segundo número termina en un dígito  
ENTONCES el subobjetivo es sumar los dos dígitos
- Prod. 2: SI el subobjetivo es sumar dos dígitos y un número es la suma de los dos dígitos  
ENTONCES el subobjetivo es extraer el número
- Prod. 3: SI el subobjetivo es extraer un número y llevamos 1 y un segundo número es la suma del primer número más 1  
ENTONCES el subobjetivo es extraer el segundo número y eliminar el que llevábamos
- Prod. 4: SI el subobjetivo es extraer un número y no llevamos nada y el número es menor de 10  
ENTONCES escribir el número y el subobjetivo es procesar los dos dígitos de la siguiente columna
- Prod. 5: SI el subobjetivo es extraer un número y no llevamos nada y el número es la suma de 10 más un dígito  
ENTONCES escribir el dígito y llevar 1 y el subobjetivo es procesar los dígitos de la siguiente columna
- Prod. 6: SI el subobjetivo es procesar los dígitos en la siguiente columna y el primer número contiene un dígito en esta columna y el segundo número contiene un dígito en esta columna  
ENTONCES el subobjetivo es sumar los dos dígitos
- Prod. 7: SI el subobjetivo es procesar los dígitos en la siguiente columna y el primer número no contiene un dígito en esta columna y el segundo dígito no contiene un dígito en esta columna  
ENTONCES el problema ha terminado

pre-establecidas que determinan el rango de situaciones ambientales a las cuales puede enfrentarse el programa y, en gran medida, también las nuevas reglas de producción resultantes (exceptuando algunos elementos aleatorios poco significativos). Pero nos encontramos, entonces, ante el habitual problema de la regresión infinita, puesto que la regla "generativa" sólo podría ser creada por el programador o, a su vez, por una regla anterior la cual, a su vez, debe haber sido generada por otra regla, etc. Nuevamente se hace patente la necesidad de recurrir, desde el enfoque computacional, al postulado innatista. Como reconocen Newell y Simon (1972) y Allport (1984), incluso este tipo de programas auto-modificables dependen, en última instancia, de reglas especificadas de forma antecedente por un programador humano. De acuerdo con Allport, estas reglas deberían representar ciertas restricciones (específicas de la especie) en el procesamiento de información; aunque no aclara si dichas restricciones deben conceptualizarse como propias de un conocimiento de tipo innato o en forma de limitaciones en el funcionamiento de la arquitectura funcional, el tipo de lenguaje utilizado sugiere que opta por esta última alternativa. Nuevamente, esto sugiere la posibilidad de situar el nivel de análisis en la arquitectura funcional del sistema; tal y como intentaremos mostrar en el capítulo 4, la estructura y funcionamiento de la arquitectura funcional (y también, en última instancia, de la arquitectura material) puede introducir restricciones en el tipo de formalización



utilizado para modelizar el sistema cognitivo humano. En particular, una arquitectura que incorpore algunos elementos de tipo continuo nos permitirá aproximarnos de forma más adecuada a ciertos hechos psicológicos ante los cuales los modelos computacionales discretos encuentran graves dificultades.

Nos parece especialmente interesante enmarcar la discusión sobre la visión computacional del aprendizaje en el campo más general de las dificultades que el enfoque computacional-proposicional encuentra para aproximarse a la conducta adaptativa de los organismos. Como señalan Carello, Turvey, Kugler y Shaw (1982), los modelos que parten de sistemas simbólicos y procesos de tipo discreto limitan de forma radical las posibilidades de auto-complicación del sistema. Más concretamente, un modelo computacional-discreto de los procesos cognitivos excluye la generación intrínseca de nuevos primitivos. En otras palabras, y como indica Rivière (1986), cualquier sistema cuyo nivel actual de competencia esté definido por una cierta lógica (la cual, obviamente, determina el poder representacional del sistema), no puede llegar a un nivel más alto de competencia mientras se mantenga en el ámbito de la realización de operaciones de modo discreto. En un sentido muy real, el "lenguaje de la mente" no puede ser modificado, puesto que su transformación exigiría, como hemos venido señalando repetidamente, la existencia de algún otro lenguaje previo que pueda expresar los predicados del nuevo lenguaje y las reglas que determinan su

extensión. En otras palabras, las operaciones que puede realizar el sistema cognitivo y los criterios de formación de sus representaciones están prefijados y no admiten modificación.

¿Qué decir, entonces, de la conducta manifiestamente adaptativa de los organismos?. Cualquier organismo se enfrenta a un medio cambiante que posee niveles de complejidad continuamente variables (especialmente en el caso de los sujetos humanos). Si, como señala Rivière (1984, 1986) el sistema cognitivo carece de mecanismos de auto-complicación, entonces debe recurrirse a ampliar el ámbito de lo innato hasta niveles difícilmente verosímiles. En otras palabras, el sistema debe tener previstas de antemano una enorme cantidad de posibilidades de complicación del medio y de respuestas ante el mismo, lo cual parece poco creíble.

Las consecuencias de la imposibilidad de aprendizaje del lenguaje de la mente pueden expresarse también en los siguientes términos: Si se acepta el postulado innatista (y no parecen quedar muchas alternativas), debemos admitir entonces que el niño posee, al nacer, las capacidades inferenciales propias de alguno de los modelos proposicionales propuestos. Por ejemplo, debemos aceptar que el niño posee innatamente la competencia (equivalente a la de la máquina de Turing) del modelo ACT de Anderson (1976). En caso contrario, estamos obligados a postular que esas capacidades se adquieren, al menos en gran parte; pero, ¿cómo puede producirse esa adquisición si esas capacidades deben

ser expresadas en forma de nuevos predicados, y éstos no pueden adquirirse sin la presencia previa de un lenguaje de la mente?. Parece obvio que este tipo de razonamientos no conduce a ninguna parte, y ello se traduce en una aceptación (más o menos explícita según el nivel de reflexión de los distintos autores) del postulado innatista, y en el escasísimo interés de los teóricos proposicionales hacia los estudios de tipo evolutivo, lo cual, como señala penetrantemente Rivière, significa renunciar a un nuevo elemento de restricción teórica y acentuar el enorme nivel de libertad formalizadora que ya poseen los modelos proposicionales. De hecho, el abandono de las explicaciones de tipo macrogenético (por ejemplo, en la línea de Piaget) por ciertas explicaciones de carácter microgenético basadas en operaciones computacionales realizadas sobre representaciones analíticas constituye otro de los caracteres restrictivos propios de las explicaciones dominantes en el enfoque computacional-discreto (Inhelder y Piaget, 1979).

#### 2.4 Redes semánticas asociativas

Uno de los aspectos más relevantes del desarrollo de los modelos computacionales en psicología ha sido su análisis del problema del significado. Se han propuesto un buen número de modelos para el almacenamiento y tratamiento de la información semántica (Quillian, 1967, 1969; Collins y Quillian, 1969; Clark, 1969a, 1969b, 1974; Schank, 1973, 1975a, 1975b, 1979, 1981, 1982; Anderson y Bower, 1973; Lindsay, Norman y

Rumelhart, 1975; Glass y Holyoak, 1974, 1975; Collins y Loftus, 1975; Anderson, 1976, 1980, 1983; Kintsch, 1974, etc.), pero prácticamente todos ellos poseen en común el hecho de formalizar la representación del significado en términos de **redes semánticas asociativas**, definidas como un conjunto de relaciones etiquetadas entre nódulos. Las redes semánticas asociativas han sido propuestas por los teóricos proposicionales como modelos para la ejecución en tareas bastante diversas (por ejemplo, comprensión de prosa, aprendizaje verbal, tareas de memoria, etc.). Este hecho no resulta en modo alguno sorprendente. Por una parte, la estructura reticular parece especialmente adecuada para expresar muchos aspectos del formalismo proposicional. Es posible, por ejemplo, representar de modo adecuado la distinción entre los elementos básicos de una proposición (por un lado sujeto y objeto, representados por nódulos; por otra parte, la relación entre ambos) en un formato reticular. Por otra parte, si se postula que el sistema cognitivo humano funciona como un sistema de símbolos discretos, resulta necesario representar también las relaciones entre conceptos en forma de cadenas de símbolos; en un formato reticular estas cadenas de símbolos corresponden a las "etiquetas" que acompañan a las relaciones asociativas entre nódulos, e introducen un elemento fundamental de diferenciación entre este tipo de planteamiento y las asociaciones inestructuradas propias de otros modelos. La necesidad de pasar de tales asociaciones inestructuradas a relaciones con un mayor nivel de

especificidad ya fue señalada clásicamente por Selz aunque, por supuesto, en un lenguaje mucho más impreciso que el propio de los actuales modelos computacionales.

El concepto de asociación etiquetada introduce, por una parte, la posibilidad de dotar a este tipo de modelos de un carácter jerárquico y, por otra, los hace especialmente importantes para la investigación en inteligencia artificial, puesto que constituyen la forma más adecuada de que los computadores digitales puedan enfrentarse a la tarea de procesar el lenguaje. En efecto, los procesos interpretativos que deben ser ejecutados sobre la red pueden tomar la forma de un programa de ordenador que tome en consideración las etiquetas asociadas a cada uno de los enlaces, de manera que pueda determinar entonces, por ejemplo, el carácter supra o sub-ordenado de un nódulo respecto a otro, las relaciones que pueden dar lugar a enunciados gramaticalmente correctos, etc. La noción de asociación etiquetada se encuentra incluso en la base de ciertos lenguajes de programación especialmente ideados para el ámbito de la inteligencia artificial (IPL, LISP, etc.).

Por su parte, la introducción de una estructura jerárquica en la red constituye un elemento fundamental de economía representativa, puesto que, como señala Quillian (1968), evita la asociación de los caracteres generales de un concepto a cada uno de sus ejemplos. Además, la existencia de este tipo de estructura parece necesaria para dar cuenta de



un conjunto bastante amplio de datos empíricos (Cofer, 1979).

Naturalmente, la aproximación al significado a partir de modelos reticulares asociativos no está exenta de problemas. Resulta especialmente interesante resaltar que muchos de ellos se relacionan, una vez más, con el excesivo poder del formalismo proposicional. Como señalan Johnson-Laird, Hermann y Chaffin (1984), un teórico proposicionalista se enfrenta con la necesidad de tomar un conjunto de decisiones sobre el modelo que está construyendo. En primer lugar, deberá delimitar el objetivo concreto de su trabajo, puesto que éste determinará algunas de las características del propio modelo. Si se pretende, por ejemplo, construir un modelo para la comprensión y producción de lenguaje, deberán tomarse decisiones en cuanto al subconjunto del lenguaje que se desee representar (por ejemplo, sobre la admisión o no de asertos que incluyan cuantificaciones). El "poder expresivo" del modelo puede radicar en mayor o menor medida en la propia red o en los procesos interpretativos definidos sobre ella: por ejemplo, Anderson (1976) sitúa la base de la interpretación de los cuantificadores en los procesos que operan sobre la red, mientras que Hendrix (1979), representa los asertos cuantificados de un modo más directo. Deberá especificarse el tipo de configuraciones que pueden ocurrir en la red: por ejemplo, algunos sistemas solamente admiten una división binaria de los enlaces que parten de cada nódulo (Anderson y Bower, 1973), mientras que otros permiten la

conexión de un número arbitrario de enlaces (Lindsay y Norman, 1977). Pueden utilizarse etiquetas sobre nódulos y enlaces (Quillian, 1968) o solamente sobre enlaces (Anderson, 1976). Todos los enlaces pueden tener la misma fuerza semántica o, por el contrario, pueden ser semánticamente heterogéneos. El conjunto de etiquetas asociativas utilizadas es bastante distinto en los diversos modelos. Las inferencias pueden ser realizadas por descomposición en primitivos semánticos en la propia red (Lindsay y Norman, 1977) o por reglas de inferencia que operan sobre ella (Anderson, 1976). Por otra parte, ¿cual es la naturaleza de esos primitivos semánticos? o, dicho de otra forma, ¿cual es el nivel de abstracción de las proposiciones?. ¿Existe distinción entre conocimiento declarativo y conocimiento procedimental?.

Como señala muy acertadamente Rivière (1986), un somero análisis de varios modelos de redes semánticas asociativas revelan una extraordinaria heterogeneidad en las respuestas que los diferentes teóricos dan a las cuestiones planteadas. Llevando el argumento al extremo, Rivière muestra las grandes divergencias de tres modelos concretos (el modelo de Clark, el de Kintsch y el sistema HAM de Anderson y Bower), incluso en el análisis de una frase bastante simple.

Resulta evidente, a partir de este análisis, que las posibilidades de restricción en este tipo de formalismos son pocas y débiles. Un modelo concreto puede, con frecuencia, permitir la derivación de predicciones empíricas, pero siempre es

posible, ante resultados adversos, acudir a un nuevo modelo formulado a partir de los mismos presupuestos. Si las redes pueden computar cualquier función, entonces resultan no ser otra cosa que formalismos vacíos, sin ningún contenido específicamente psicológico (Paivio, 1977; Hayes-Roth, 1979; Lachman y Lachman, 1979; de Vega, 1981).

El tratamiento que las redes semánticas dan al significado participa del carácter circular que hemos atribuido a los modelos proposicionales en general. El significado de un concepto consiste en su relación con otros nodos conceptuales los cuales, a su vez, adquieren su significado en virtud de su relación con otra serie de nodos (entre los que se incluyen los que mencionábamos en primer lugar), etc. Esto es lo que Anderson (1980) denomina significado configuracional. Pero, obviamente, nos encontramos ante un sistema de referencias circulares; y si, como pretenden algunos teóricos proposicionales, deseamos romper esta circularidad postulando la existencia de algunos elementos terminales de carácter sensorial (Anderson, 1980), chocamos frontalmente con otra dificultad, puesto que, como se ha señalado repetidamente, desde un enfoque computacional el resultado del análisis perceptivo (del transductor pylyshyano) no es otra cosa que una cadena de símbolos discretos cuya interpretación nos remite, una vez más, a la red asociativa cuyos referentes externos pretendíamos infructuosamente descubrir. Dicho de otra forma, los modelos de red semántica presentan un análisis de las relaciones

intensionales entre los conceptos (el "sentido" fregeano); la representación y evaluación de estas relaciones intensionales pueden (en realidad, **deben**) realizarse independientemente de las relaciones extensionales de los mismos conceptos (el "referente" de Frege).

Consideremos ahora algunos aspectos del lenguaje que los modelos de redes semánticas no parecen capaces de abordar satisfactoriamente. En particular, el lenguaje natural se caracteriza por su ambigüedad a nivel léxico, estructural, semántico y pragmático, lo cual ha frustrado en gran parte las esperanzas de conseguir programas de traducción mecánica o de comprensión automática del lenguaje que puedan operar de forma eficaz sin un soporte humano (Winograd, 1984).

Un somero repaso a los distintos niveles de análisis necesarios para la comprensión del lenguaje (figura 2.3) muestra rápidamente las dificultades que encuentran los ordenadores (que, recordemos, trabajan casi siempre en base a redes semánticas del tipo que estamos analizando) para abordar la cuestión. Normalmente los computadores trabajan con lenguaje "escrito" en la pantalla, lo cual permite evitar las dificultades asociadas al análisis fonético del mensaje y concentrarse de inmediato en su análisis morfológico. Señalemos, además, que este tipo de entrada de información evita (pero, por supuesto, no soluciona) el problema absolutamente esencial de la traducción de una imagen audio-verbal a una cadena de símbolos discretos, es decir, ignora el problema de definir el transductor de Pylyshyn.

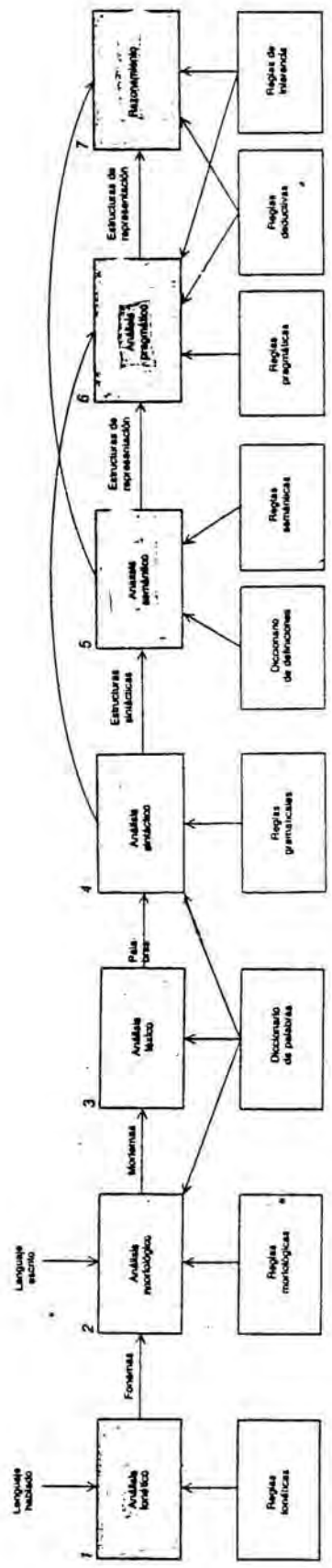


Figura 2.3 Aspectos básicos del procesamiento del lenguaje



Para la realización de un análisis morfológico el programa necesita, en primer lugar, una serie de reglas que le permitan descomponer una palabra en radical y flexiones, y otras reglas que asocien, por ejemplo, las diversas flexiones de una forma verbal con los distintos tiempos. Es necesario almacenar también un conjunto de excepciones significativas en cuanto, por ejemplo, a verbos irregulares.

Para cada radical que resulta del análisis morfológico, un "diccionario" permite situar cada palabra en su categoría léxica (aunque algunos lexemas pueden pertenecer a más de una categoría). En este nivel se presenta la ambigüedad léxica, en la cual una misma palabra puede tener más de un sentido. La forma más habitual de soslayar esta dificultad consiste en un análisis estadístico del entorno de la palabra, de forma que la interpretación de ésta dependerá del contexto en el que se encuentre. Sin embargo, en la práctica este procedimiento produce un número de errores muy considerable.

En el caso de la ambigüedad estructural, el problema trasciende los términos individuales y se refiere a la interpretación global de las frases. El traductor se ve con frecuencia obligado a escoger entre dos estructuras gramaticales igualmente verosímiles. En algunos casos, además, puede presentarse una ambigüedad de "estructura profunda"; se trataría del caso en que dos posibles lecturas de una frase presenten la misma estructura aparente y diferir, en cambio, en el significado. Es el caso de la frase "Los pollos están

listos para comer"; pero, ¿los pollos comerán o serán comidos?. El formalismo ideado para representar la estructura profunda del lenguaje ayuda muy poco a determinar la estructura profunda "verdadera" de una frase ambigua (Winograd, 1984).

Puede hablarse también de ambigüedad semántica. Este tipo de ambigüedad se presentaría cuando una frase puede tener distintas funciones en el sentido global de una oración. Por ejemplo, una parte de cierta oración puede poseer un carácter referencial, o ser más bien atributiva. Cualquier procesador de lenguaje deberá ser capaz de enfrentarse también a este tipo de situaciones.

Finalmente, puede mencionarse el caso de la ambigüedad pragmática, que surge de los problemas de traducción de pronombres y otras palabras particulares. Por ejemplo, si un ordenador que base su funcionamiento en redes semánticas asociativas intenta traducir al castellano la frase "She dropped a plate on the table and broke it" se encontrará inmediatamente con el problema del género que debe atribuir a la traducción de "it". Esta elección implica suponer que lo roto ha sido el plato, o bien la mesa. Como en los casos anteriores, los programas de ordenador contruídos hasta el momento han sido incapaces de resolver este tipo de complicaciones.

Los problemas de las redes semánticas del tipo que estamos analizando no se reducen a la cuestión de la ambigüedad, aunque ésta constituya una dificultad fundamental. Por citar solamente una dificultad específica, nos referiremos

brevemente al problema de la representación de relaciones como la sinonimia y la antonimia. Hermann, Chaffin y Daniel (1984) demuestran, por ejemplo, que los juicios sobre el carácter antónimo de dos palabras son más rápidos si los dos elementos se encuentran opuestos simétricamente. ¿Cómo se expresa la simetría en el tipo de estructuras que estamos analizando?. Parece bastante evidente la necesidad de retornar a una formulación de carácter más contínuo que nos permita hablar de "distancias" entre elementos e introducir propiedades "espaciales" o "geométricas" como la simetría.

Otro tipo de dificultades se refiere a los procesos inferenciales que deben realizar este tipo de redes. Citemos únicamente un ejemplo problemático: A está sentado a la derecha de B; B está sentado a la derecha de C. ¿Implica esto que A está sentado a la derecha de C?. Una respuesta adecuada a esta pregunta exige como requisito imprescindible la referencia a ciertas propiedades del mundo externo. En particular, en este caso será necesario determinar si los sujetos están sentados ante una mesa rectangular y mirando en la misma dirección, o bien si se encuentran situados en torno a una mesa de forma circular. En otras palabras, este tipo de cuestiones requiere un sistema inferencial que tenga acceso a la representación de la situación a la cual se hace referencia.

Los modelos de redes semánticas son particularmente vulnerables a la crítica en términos de lo que se ha denominado "falacia simbólica". Los teóricos que trabajan en este

ámbito parecen suponer que el proceso de trasladar el lenguaje natural a este tipo de redes hace posible su comprensión. Sin embargo, en ocasiones parece olvidarse que una representación sólo adquiere significación, en sentido estricto, si está conectada de alguna forma con el mundo externo. El solipsismo propio de los planteamiento que estamos analizando se concreta en la "falacia simbólica": de acuerdo con esta suposición, la mera traslación de las sentencias del lenguaje natural a cadenas simbólicas constituiría una aproximación a su significado. El problema es muy semejante al que se ha planteado en ocasiones al criticar la idea de que un ordenador pueda poseer "estados simbólicos". Como señala muy acertadamente Heil (1980), los estados internos de una máquina de computación pueden ser tratados como simbólicos solamente a causa de que están conectados a las actividades representacionales de agentes (los sujetos humanos) que dan sentido a su output. Del mismo modo, si consideramos el sistema cognitivo humano como un mecanismo computacional, es necesario afirmar con toda rotundidad que la atribución de estados simbólicos puede resultar absolutamente artificial. Como se ha dicho en algunas ocasiones, resulta totalmente erróneo atribuir al sujeto lo que no son sino representaciones del psicólogo. Por otra parte, ¿de qué modo puede una maquinaria solipsista -el sistema cognitivo de un sujeto- atribuir estados simbólicos a otra maquinaria solipsista -otro sujeto o un ordenador- sin referirse en algún momento a la realidad externa?. Y, en cuanto intenta

remitirse a esa realidad externa (supuesta), el psicólogo de base computacional debe relajar necesariamente algunas de sus fuertes presuposiciones y, en particular, se ve obligado a aceptar la existencia de códigos representacionales distintos al formalismo proposicional (ver apartados anteriores). A nuestro entender, y en pleno acuerdo con la formulación propuesta por Rivière (1986), las convenciones comunicativas y, muy particularmente, el lenguaje verbal, pueden encontrar su origen en referencias compartidas basadas en representaciones que originalmente poseen un carácter analógico y continuo. Esta nos parece la única alternativa válida a la posición innatista adoptada explícita o implícitamente por los teóricos computacionales.



3. El cambio desde la perspectiva  
general de la ciencia

---

En una visión globalizadora, la ciencia moderna (o, al menos, una parte de ella) nos presenta la naturaleza como un conglomerado de materia, energía e información (Campbell, 1983). La compleja relación entre estos componentes fundamentales puede ser puesta de manifiesto en numerosos ejemplos. Por citar solamente uno, Bennett y Landauer (1985) presentan un brillante análisis sobre las limitaciones físicas de los procesos de computación y muestran, por ejemplo, que no existe un mínimo de energía necesario para la realización de un determinado cómputo, tanto si el análisis se realiza desde la termodinámica clásica como desde la mecánica cuántica. Por supuesto, la utilización de cantidades de energía cada vez menores tiene el precio de aumentar el tiempo necesario para la computación.

Este tipo de estudios plantea preguntas auténticamente interesantes, del tipo: ¿exigen las operaciones elementales de procesamiento como mínimo un determinado tiempo? o, en relación con la capacidad de almacenamiento de los sistemas computacionales, ¿cual es el máximo tamaño que puede darse a un sistema de memoria?, ¿cuantas partículas del Universo pueden ser reunidas y ensambladas a efectos de almacenamiento?. Obviamente, el tamaño de la memoria de cualquier sistema de computación impone limitaciones a la precisión de sus cálculos. Si se analizan detenidamente estas cuestiones se advertirán inmediatamente algunas implicaciones profundas, derivadas del hecho de que lo que se plantea en realidad son las limitaciones que pesan sobre la ejecución material de

operaciones matemáticas y, en última instancia, sobre la misma cognoscibilidad de las leyes del Universo. Si cada cómputo exige un mínimo de tiempo, ¿existirán cálculos imposibles de realizar antes de que la maquinaria computacional se deteriore antes de quedar inservible?; y, en consecuencia, ¿será imposible sobrepasar un cierto grado de precisión en dichos cálculos?. No podemos olvidar el hecho de que las leyes físicas, en las cuales deberíamos buscar respuesta a algunas de estas cuestiones, vienen expresadas por medio de operaciones matemáticas cuya resolución, presumiblemente, quedaría limitada por la acción de esas mismas leyes físicas. Citando a Bennett y Landauer (1985), "nos estamos preguntando cual es la forma definitiva en que pueden ser aplicadas las leyes físicas, dadas las restricciones impuestas por el universo que las propias leyes describen".

En el fondo, las consideraciones anteriores no son sino una forma de expresar algunas inquietudes fundamentales de la ciencia contemporánea que, al menos en buena parte, se relacionan con la vieja idea del demonio laplaciano y, en consecuencia, con el binomio determinismo-azar. Como es sabido, Laplace imaginó la posibilidad de existencia de una inteligencia infinitamente superior a la humana en lo cuantitativo, pero fundamentalmente idéntica en lo cualitativo. A partir de una capacidad de cálculo ilimitada, el demonio de Laplace sería capaz, a partir de la aplicación de leyes físicas deterministas, de reconstruir de forma completa el pasado o de predecir infaliblemente el futuro, siempre que

conociera de forma suficientemente precisa las condiciones iniciales del sistema considerado. Resulta casi superfluo decir que el computador es una especie de encarnación moderna del demonio laplaciano, aunque sus capacidades estén todavía lejos de las requeridas para la monumental tarea que estamos considerando. El planteamiento de Bennett y Landauer conduce a pensar que estas capacidades no podrán ser alcanzadas nunca y que, en consecuencia, el determinismo absoluto no es sino una respetable noción metafísica.

El demonio laplaciano -cuya viabilidad fue puesta en cuestión por su propio creador al adoptar un enfoque probabilístico- se halla ante la dificultad fundamental de que la precisión exigida en el conocimiento de las condiciones iniciales del sistema para que su tarea sea realizable es de orden infinito. Ninguna precisión finita, por grande que sea, es suficiente para determinar la evolución posterior del sistema, puesto que el número de trayectorias evolutivas derivables de datos iniciales de precisión no infinita es infinito, aunque puedan acotarse los límites de estas trayectorias. Puede citarse como ejemplo el estudio clásico de Poincaré (1914), en el cual se demuestra que, aun conociendo las leyes de la mecánica newtoniana y los factores implicados en ellas, es imposible determinar cual será la cara que muestra un dado una vez lanzado. A partir de cualquier intervalo de valores de las condiciones iniciales, por pequeño que sea, se obtiene un número igual de trayectorias que conducen a cada uno de los posibles resultados.

Se advierte de inmediato que a la trilogía materia-energía-información subyace una visión de las relaciones entre ciencias naturales y las llamadas ciencias "humanas" que poco tiene que ver con las aproximaciones clásicas a la cuestión. La distinción entre "Naturwissenschaften" y "Geisteswissenschaften" (Windelband, 1894) parece perder, en el marco de la ciencia actual, buena parte de su sentido, pero no por reducción de la psicología a la biología o a la física, sino porque la propia evolución de estas ciencias ofrece un marco mucho más propicio a la tarea de tender puentes entre los dos ámbitos.

Tomemos como ejemplo el desarrollo de algunos aspectos de la física contemporánea. Parece extenderse, entre algunos investigadores que trabajan en el ámbito de las ciencias "humanas", quizá bajo el influjo de la teoría general de sistemas, la creencia de que pueden encontrarse puntos de contacto entre el análisis de sistemas complejos de carácter interactivo en física y el estudio de la dinámica de los procesos sociológicos, psicológicos e incluso históricos. Por su parte, la evolución de la física parece tender a dar marcha atrás en la ruptura que, en algunos momentos, ha existido entre esta ciencia y las demás disciplinas. Intentaremos analizar algunos puntos clave de esta transformación.

Prigogine y Stengers (1979) analizan los aspectos fundamentales de la aproximación clásica a la ciencia. De acuerdo con el punto de vista clásico, el universo puede ser concebido como una especie de autómeta; como indicábamos algunos



párrafos más atrás, la prescripción de unas condiciones iniciales definidas permitiría, al menos en teoría, conocer la evolución perfectamente predeterminada de todo el sistema. La posición del hombre en este tipo de universo es, de alguna forma, contradictoria: por una parte, el hombre podía, potencialmente, controlar la evolución del universo hasta sus últimos detalles. El futuro sería, por tanto, de potencial omnipotencia. Pero, por otro lado, en este universo mecánico la misma existencia de la vida debe ser considerada como una extraña casualidad, como el producto de la concatenación de una serie de sucesos bastante improbables. Se trata de la conocida tesis de Monod (1981) según la cual la existencia de vida en este tipo de universo es, obviamente, compatible con sus condiciones iniciales, pero de ninguna manera es deducible de ellas. Contrariamente, se trataría de un fenómeno cuya aparición, de acuerdo con las condiciones iniciales del sistema, resulta altamente improbable. De este modo, la existencia de seres vivos y, en particular, de la especie humana, adquiere un carácter de inverosímil casualidad, de radical extrañeza. Este es, de alguna forma, el precio a pagar si se postula el carácter completamente determinista de la evolución del universo.

Una de las concreciones más conocidas de esta visión "pesimista" de la evolución del universo es la termodinámica clásica y, muy especialmente, su segundo principio. La termodinámica resulta relevante para la biología y, al menos potencialmente, para la psicología, desde el momento en que sus

dos principios fundamentales rigen el conjunto de las transformaciones físico-químicas (¿e informacionales?) que tienen lugar en los sistemas observables. El primer principio afirma la conservación de la energía total de un sistema en el transcurso de dichas transformaciones. El segundo principio, en su versión original, describe la evolución de un sistema aislado: existirá una magnitud  $S$  que solamente puede crecer durante el desarrollo de cualquier transformación de energía, de forma que, transcurrido un tiempo suficientemente largo, alcanza un valor máximo que caracteriza al estado de equilibrio termodinámico. El segundo principio adquiere una fuerza especial con la interpretación estadística aportada por Boltzmann: la entropía  $S$  es una medida del desorden molecular. De esta forma, el segundo principio se convierte en una ley de desorganización progresiva, y condena al universo a la muerte entrópica por indiferenciación.

No resulta sorprendente, por tanto, que para un filósofo como Nietzsche la termodinámica representara el modo en que la ciencia participa en el nihilismo del pensamiento moderno. Como señala Deleuze (1973), en este sentido Nietzsche resulta ser uno de los más agudos críticos del enfoque mecanicista y de la tendencia a la indiferenciación que se deriva de la termodinámica clásica.

Resulta evidente que, si se desea ampliar de modo útil la teoría clásica del equilibrio a los sistemas vivos es necesario relajar algunas de las fuertes restricciones que tal teoría implica. En primer lugar, resulta fundamental la

observación de que los seres vivos son sistemas termodinámicamente abiertos, es decir, que intercambian materia, energía e información con su entorno. Resulta también obvio que los sistemas vivos evitan las situaciones de equilibrio termodinámico y, en particular, son capaces de mantener estados estables distintos de la "muerte calórica" (se trata de lo que denominaremos estados estacionarios). En estos casos, la entropía positiva producida en el sistema es disipada en el entorno, de forma que es posible mantener una estructura constante.

Schrödinger (1945) intentó aplicar el concepto de orden entrópico de Boltzmann para dar cuenta de la estabilidad de los sistemas vivos. De acuerdo con su interpretación, la diferenciación de los organismos se consigue disipando más entropía de la que se produce, es decir, "robando orden" al ambiente por medio de un flujo negativo de entropía, la **neguentropía**. Sin embargo, como señala Wagensberg (1985), existe la sospecha generalizada de que el concepto de orden de Boltzmann resulta insuficiente para explicar la complejidad de las funciones biológicas. Esta idea resulta especialmente importante, particularmente porque lleva implícita una crítica a la teoría de la información que, como es sabido, comparte en buena parte la interpretación boltzmanniana. Algo semejante es lo que quiere indicar Thom (1980a) en su crítica a la introducción de la teoría de la información en ciertos ámbitos de la biología y, en especial, a su aplicación a los fenómenos de transmisión hereditaria. Como

veremos algo más adelante, esta idea será útil para nuestro enfoque del significado en psicología.

Como señala Prigogine (1983a), las insuficiencias de la interpretación boltzmanniana no implican la irrelevancia de la termodinámica para el estudio de los sistemas vivos. Al contrario, de la propia termodinámica puede derivarse un análisis alternativo que presenta un conjunto de implicaciones absolutamente fundamentales. La idea básica de este enfoque es relativamente sencilla: cuando los sistemas se alejan mucho del estado de equilibrio termodinámico, su evolución deja de describirse en términos de un régimen lineal para entrar en el no lineal. En este régimen aparecen discontinuidades e inestabilidades, de forma que el estado estacionario compatible con las condiciones deja de ser único. Se abre así el camino para un análisis de la evolución de los sistemas vivos (y también de muchos sistemas no vivos) en términos de no linealidad, fluctuaciones, bifurcaciones, etc. En definitiva, se trata de un planteamiento que representa, en buena medida, una ruptura con el enfoque determinista y mecanicista clásico. Esta reconceptualización de las ciencias físicas es descrito por Prigogine (1983a) como un rechazo a la búsqueda de la simplicidad que se encuentra en la obra de autores tan fundamentales como Galileo, Descartes y Newton. Se trata, por el contrario, de plantear una aproximación a la complejidad.

Permítasenos realizar aquí un inciso. Algunos psicólogos, y muy en particular Kurt Lewin, han valorado muy positivamente

el enfoque científico galileano. Para Lewin (1935), es factible superar la distinción entre los enfoques ideográfico y nomotético mediante una aproximación adecuada a la construcción de teorías. A partir de las leyes generales, debe ser posible representar situaciones concretas de tal forma que el hecho observado pueda derivarse de ellas según los principios dados en aquellas leyes. En otras palabras, si se describe una situación mediante un conjunto de factores "condicional-genéticos" o genotípicos (propiedades conceptuales) el curso de los sucesos observables ("fenotipo") puede deducirse de esta descripción como una consecuencia lógica: es decir, la representación correcta de lo que "es" constituye al mismo tiempo una aplicación de lo que sucede (Wolman, 1960).

Esta tendencia al análisis simultáneo de las leyes generales y de los casos específicos distingue, de acuerdo con Lewin, las dos grandes etapas de la psicología: por una parte, la psicología "especulativa" o aristotélica, una "teoría de la clase" en la cual la pertenencia de un objeto a cierta clase determina su esencia y, por tanto, la totalidad de su comportamiento; desde este enfoque, uno de los objetivos primordiales de la ciencia y de la filosofía sería el de hallar el conjunto de predicados atribuibles de forma común a un conjunto de objetos, que quedarían definidos entonces como clase. Este punto de vista centrado en el "fenotipo" habría sido substituído, de acuerdo con Lewin, por el énfasis galileano hacia los conceptos condicional-genéticos propios



de un "genotipo". Sin embargo, y aun aceptando la verosimilitud de sus argumentos, Lewin parece olvidar, en su valoración globalmente positiva de la herencia galileana, el hecho de que algunas de las concepciones del científico de Pisa resultan decisivas en la construcción de la cosmovisión newtoniana. Nos referimos, en particular, al análisis galileano del movimiento, en el cual el énfasis en lograr una equivalencia cuantitativa entre la "causa" (el peso) y el "efecto" (el movimiento) lleva a considerar éste como un simple estado y a centrar el interés en la aceleración, que sería constante al igual que el peso. De esta forma, lo que es esencialmente variación queda reducido a constancia.

En la concepción newtoniana la física se ocupa esencialmente de procesos que poseen un carácter determinista y reversible. Puede citarse como caso paradigmático el estudio del movimiento: definida la función que describe la trayectoria de un móvil, resulta posible "avanzar" o "retroceder" a lo largo del tiempo, en el sentido de que puede determinarse perfectamente la posición del móvil en un momento  $t$  cualquiera. Se trata, en definitiva, de una concreción parcial y limitada del ideal laplaciano: dado un momento  $t$  y una función  $f$ , podemos determinar por aplicación de esta función el estado del sistema en el momento  $t$ .

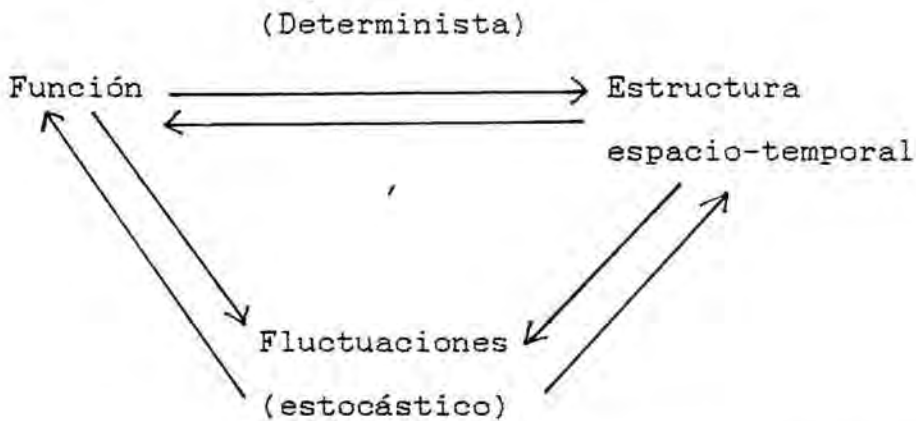
El problema se plantea cuando, a la luz de la física contemporánea, empiezan a introducirse correcciones importantes a algunos de los fundamentos de este punto de vista. Por ejemplo, las partículas elementales, lejos de constituir el

soporte permanente e invariable de las apariencias macroscópicas cambiantes, han resultado ser muy poco estables (salvo excepciones). De este modo, una concepción que se remonta a los atomistas griegos pierde una buena parte de sus fundamentos. Otro ejemplo: el estudio de los fenómenos macroscópicos complejos (más allá de las "trayectorias" newtonianas) revelan cada vez con mayor claridad la existencia de lo casual y lo irreversible, es decir, de fenómenos que en la concepción clásica estaban considerados, en el mejor de los casos, como excepcionales. Al contrario, desde la perspectiva actual es posible afirmar que los procesos reversibles y deterministas propios de la concepción newtoniana resultan ser, en muchos casos, idealizaciones bastante artificiosas. Los conceptos básicos del nuevo enfoque que estamos considerando son, por tanto, los siguientes: El no equilibrio (o el alejamiento suficiente del equilibrio) es una fuente de un nuevo tipo de orden, denominado orden por fluctuaciones, que implica un cambio en la organización del sistema y permite, además, la presencia de un comportamiento coordinado y "cooperativo" entre un conjunto de unidades que, en situaciones próximas al equilibrio, no mantenían relación entre sí. En estados próximos al equilibrio se producen ciertas fluctuaciones aleatorias; por ejemplo, en el caso de un sistema químico las fluctuaciones locales en las concentraciones son debidas al movimiento molecular. En un ecosistema, las fluctuaciones pueden ser debidas a transformaciones en el medio referidas a la cantidad de recursos disponibles. En

condiciones normales, estas fluctuaciones no tienen posibilidad de ampliarse y prosperar; normalmente, tenderán a "regresar", es decir, a retornar a un estado cercano al de equilibrio. Pero, superada una distancia crítica al estado de equilibrio, estas fluctuaciones espontáneas pueden amplificarse y arrastrar al sistema hacia nuevos estados estables: las estructuras disipativas. La idea de estructura disipativa es la de un régimen estable que se mantiene por intercambio de materia y energía (y, posiblemente, información) con el medio. Esto es lo que diferencia, por ejemplo, el tipo de estabilidad propio de los sistemas vivos del que está presente en estructuras no vivas organizadas, como los cristales. En este último caso, el sistema puede permanecer aislado sin que por ello deba producirse la transición hacia el estado de equilibrio termodinámico. En el caso de los organismos vivos, el aislamiento no puede conducir sino a la desorganización.

De acuerdo con Prigogine (1977, 1983b), la evolución de un sistema cualquiera se desglosa en dos fases. En primer lugar, el régimen existente entre las inestabilidades, que es determinante en el sentido de que ciertas ecuaciones, como las de cinética química o dinámica de población (o de cambios de activación en sistemas distribucionales, añadimos nosotros) determinan los cambios que se producen en las variables del sistema. Lejos del equilibrio, la evolución del sistema presenta un elemento de aleatoriedad, puesto que los cambios que en él se produzcan vendrán determinados por la

primera fluctuación que tenga lugar. El comportamiento del sistema puede resumirse entonces en el siguiente esquema:



Podemos considerar la función como la "micro-estructura" del sistema, mientras que la organización a gran escala espacio-temporal corresponde a la "macro-estructura". Una fluctuación da lugar a una modificación local de la microestructura que, en caso de que los mecanismos reguladores sean insuficientes, modificará la macro-estructura.

¿Hasta qué punto es aplicable este punto de vista a ámbitos distintos al de la física o la química?. Podemos pensar inmediatamente en algo como la teoría de sistemas, aunque quizá sea necesario introducir algunas precisiones. Resulta fácil encontrar ejemplos, en la biología o en la sociología, que presenten analogías con el tipo de hechos fisico-químicos que estamos considerando. Centrémonos, por ejemplo, en el caso de los eco-sistemas. y consideremos una de las aportaciones fundamentales de Margalef (1968, 1972),

a saber, la introducción de la teoría de la información en la dinámica de poblaciones:

Supongamos que se utiliza la medida de información propuesta por Shannon para evaluar la diversidad de los ecosistemas. La idea básica consiste en que la probabilidad  $P_i$  presente en la fórmula fundamental de la teoría se refiera a la probabilidad de que un individuo cualquiera pertenezca a la especie  $i$ . De esta forma, la entropía sería una medida de lo que Margalef llama el "barroquismo" del sistema. Lo fundamental es que el índice de complejidad necesario para que el sistema sea estable se encuentra siempre en un intervalo bien definido. ¿Se puede encontrar un fundamento físico que explica la acomodación estructural de los eco-sistemas?. Lurié y Wagensberg (1983) responden afirmativamente en base a la aplicación de un principio variacional.

Sin embargo, el ejemplo anterior parece poco satisfactorio, puesto que parece implicar, de alguna forma, un reduccionismo fisicalista difícil de admitir. No se trata ahora de encontrar un fundamento físico que dé cuenta de cualquier morfología en la que se produzcan cambios repentinos, ni de buscar la base de esos cambios en la interacción de partículas físicas, sino de construir una teoría general de la estabilidad estructural aplicable a todo sistema que cumpla con un conjunto de requisitos mínimos. Prigogine (1983b) afirma que esta tarea es posible y sugiere cuales deben ser las características de un sistema al cual sean aplicables los conceptos que estamos analizando:



1) Debe tratarse de sistemas caracterizados por un comportamiento coherente de un cierto número de unidades. En este contexto, comportamiento coherente implica cooperación entre esas unidades. En el ámbito biológico, podemos citar de forma inmediata multitud de ejemplos de este tipo de sistemas: el más sencillo es el de un organismo unicelular, que utiliza en su metabolismo millares de enzimas. En un terreno más interesante para nosotros, es casi innecesario señalar el ejemplo del sistema nervioso, y particularmente del cerebro, donde millones de neuronas presentan un comportamiento coherente. Por citar un caso concreto, desde el punto de vista del comportamiento global está establecido desde hace tiempo que el sistema nervioso central es sede de fenómenos rítmicos persistentes con frecuencias características reproducibles (Boyarski, 1967). Wilson y Cowan (1971) han presentado las ecuaciones diferenciales no lineales que describen la dinámica y los acoplamientos entre poblaciones neuronales localizadas y demuestran que este sistema puede presentar comportamientos rítmicos cuya frecuencia fundamental depende del estímulo aplicado. Existen también estudios sobre el proceso de despolarización de la membrana neuronal que muestran que la diferencia de concentración iónica entre los dos lados de la membrana juega un papel restrictivo que mantiene al sistema lejos del equilibrio; en estas condiciones puede producirse una inestabilidad que genere un estado de despolarización cíclica (Blumenthal, Changeix y Lefever, 1970).

Este tipo de análisis ha sido aplicado a ámbitos tan dispares como el urbanismo, la construcción de modelos "dinámicos" en economía en oposición a los modelos lagrangianos, el comportamiento coherente de poblaciones de insectos, etc.

2) El problema de los límites del sistema desempeña un papel importante. Se trata siempre de sistemas no aislados, que intercambian energía, materia e información con el mundo externo.

3) La evolución temporal de estos sistemas muestra, incluso en intervalos muy cortos, puntos singulares que marcan la aparición de fenómenos macroscópicos nuevos y "repentinos", no previsibles completamente a partir de la evolución anterior del sistema.

En todo caso, pensamos que la introducción de este tipo de nociones en el ámbito de la psicología debe realizarse con precaución, adjudicándole de momento el carácter de analogía potente y relevante, pero evitando cualquier tentación reduccionista. En este sentido, nos parece interesante una aproximación matemática como la propuesta por la teoría de las catástrofes, puesto que la posibilidad de modelizar las discontinuidades de diversos sistemas mediante instrumentos matemáticos comunes puede introducir elementos de valoración de la analogía.

Entraremos ahora en un aspecto bastante distinto de la problemática que estamos analizando. Se tratará, a partir de ahora, de valorar el paralelismo entre algunos problemas planteados en el campo de la biología (y posiblemente de

algunas de sus soluciones) y las cuestiones básicas planteadas para los psicólogos cognitivos.

Por ejemplo, el enfoque computacional que hemos analizado en capítulos anteriores está también presente en el ámbito de la biología. En general, puede afirmarse que los dos tipos de computadores actualmente disponibles, analógico y digital, reflejan con notable fidelidad los dos enfoques, profundamente distintos, con los que se aborda el estudio del sistema cognitivo humano. El enfoque analógico destaca los aspectos dinámicos, en buena parte continuos, mientras que el enfoque digital pone énfasis en aspectos discontinuos, "lógicos" y cuasi-estáticos. Algo muy semejante puede afirmarse del ámbito de la biología (Goodwin, 1976; Carello et al., 1982). En muchas áreas de la biología (en particular, en aquellas en las que la bioquímica juega un papel central) el origen de las discontinuidades es generalmente explicado por la existencia de un umbral. Tomando el ejemplo propuesto por Goodwin (1976), la diferenciación de un organismo simple en regiones distintas (cabeza, tronco...) es explicada, generalmente, mediante el supuesto de que existe un gradiente de concentración continuo de alguna substancia a lo largo de todo el eje de simetría del organismo, y un "detector" de umbral en cada una de las células. Todas las células que detectan un nivel de concentración superior al umbral se diferencian, por ejemplo, en células de la cabeza, mientras que las que se encuentran por debajo de ese umbral se convierten en células del tronco. Este tipo de mecanismos es

utilizado también para explicar las "decisiones" tomadas por las células en el curso de su ciclo de crecimiento y división. Obviamente, podemos centrar nuestro análisis en un nivel macro-molecular en el cual el comportamiento discontinuo de los agregados moleculares nos permitirá tratar a la célula como un tipo particular de máquina de Turing (Stahl, 1965). Sin embargo, los problemas de este enfoque comienzan a hacerse evidentes de forma casi inmediata. Por ejemplo, el propio Stahl (1966) muestra que este tipo de "células-autómata" es incapaz de mostrar una evolución de tipo lamarckiano.

La secuencia de acontecimientos que constituyen el ciclo celular en un modelo de tipo discreto es de un tipo "causal simple", en el sentido de que se modeliza como un conjunto de "decisiones" que determinan la lectura serial de los genes. Desde este punto de vista, se ignoran los problemas relativos a la interacción entre genes, regulación de su expresión, variaciones del proceso como consecuencia de las condiciones ambientales, etc. Cualquier acercamiento a este tipo de fenómenos parece exigir previamente la aceptación de algún tipo de proceso regulador de base continua. En esta línea, la evidencia existente es perfectamente concluyente en su demostración de que los genes (al menos, los estudiados hasta el momento, como los responsables de la producción de fosfatasa alcalina y galactosidasa) son regulados de un modo continuo por encima de un orden de magnitud mediante la variación continua del nivel represor (es decir, del umbral

a partir del cual la presencia de concentraciones mayores de sustancia represora impedirían la expresión del gen). De este modo, el problema de la regulación de los genes pasa a ser más una cuestión de la "teoría de control" (basada en procesos de tipo continuo) que de la teoría de autómatas. Es en este sentido en el que Rivière (1984, 1986) puede afirmar que "el significado del código genético no puede establecerse en términos simples de traducción entre códigos discretos" (1986, pág. 54), puesto que se hace imprescindible la referencia a mecanismos regulatorios de tipo continuo. Como señala este mismo autor, "el significado... está precisamente en el proceso dinámico y continuo de despliegue, que no puede describirse de forma completa mediante la especificación de los códigos discretos en la célula" (1986, pág. 54).

Otro aspecto interesante se encuentra en la respuesta que Gregory (1976) proporciona a la relevante pregunta: ¿cómo es posible que tan poca información (procedente del entorno) controle tanto comportamiento?. Se advierte inmediatamente el paralelismo de este planteamiento con el que realizan los teóricos computacionales. Pero la respuesta de Gregory es bastante distinta. Efectivamente, se acepta que la información evoca modelos pre-existentes de objetos del mundo perceptible que han sido formados y almacenados en el cerebro. Pero, a diferencia de lo que postula el enfoque computacional-discreto, Gregory sugiere que estos modelos poseen un carácter analógico. Ciertamente, un computador continuo que funcione directamente a partir de entradas



variables carecerá de un gran poder de análisis pero, en contrapartida, pueden operar tan rápidamente como lo exija la sucesión de cambios en sus entradas. Es, por tanto, ideal para la computación en tiempo real y, al menos en el ámbito perceptual, no necesita la precisión o el poder analítico de un sistema digital que siga las etapas simbólicas de un tratamiento matemático de la situación. La conclusión de Gregory es que un sistema analógico continuo es apropiado para el procesamiento de datos perceptuales. Como veremos en el siguiente capítulo, sostendremos que este tipo de sistemas es también adecuado para el procesamiento de tipo más "central".

En su comentario sobre la cuestión anterior, Waddington (1976) sugiere que el término "modelo" puede ser substituído por el de "creodo". Entonces, sea  $X$  un conjunto de estímulos externos. Este conjunto es aplicado en el conjunto de los estados cerebrales  $Y$  mediante una cierta función dependiente de las conexiones neuronales entre el cerebro y los órganos de los sentidos. La idea es que el espacio de los estados cerebrales está dividido en un cierto número de dominios, cada uno de los cuales está caracterizado por un campo de vectores que, en un instante dado, es dominado por un atractor particular. Sin entrar en detalles, este planteamiento pertenece completamente al ámbito de la teoría de las catástrofes, un instrumento matemático cuya utilización para la modelización de procesos psicológicos será defendida en los siguientes capítulos.

Para no alargar la exposición, nos limitaremos a citar otros desarrollos de la biología que nos parecen interesantes para la problemática psicológica. Pueden mencionarse, por ejemplo, los estudios sobre las bases físicas de la codificación (con interesantes referencias a la relevancia del nivel cuántico), la teoría dinámica de la morfogénesis (derivada directamente de la teoría de las catástrofes), el análisis de la noción de orden (que incide en algunos aspectos que hemos mencionado anteriormente), el estudio de los autómatas autoreproducibles, construcción aleatoria de redes (ver capítulo siguiente), el análisis del funcionamiento cerebral en términos de espacios de tolerancia, etc. Sin necesidad de profundizar en el análisis de estos aspectos (lo cual podría ser el objeto de una Tesis completa), resulta evidente que la problemática planteada a la biología presenta puntos de contacto fundamentales con los grandes temas de la psicología (y, en particular, de la psicología cognitiva); en alguna medida, las soluciones planteadas por los biólogos a este tipo de cuestiones pueden ser un punto de referencia importante para el trabajo de investigación en psicología. Y, también en buena medida, las respuestas planteadas por los biólogos no parecen apoyar la pretensión de abordar el estudio del comportamiento exclusivamente en términos de sistemas digitales basados en operaciones computacionales de carácter discreto.

4. De la máquina de Turing a la  
máquina de Boltzmann

---

#### 4.1 Modelos distribucionales interactivos

En los capítulos anteriores hemos proporcionado argumentos en favor de un análisis micro-estructural de los procesos cognitivos y, en particular, hemos sugerido la necesidad de valorar la utilidad de modelos micro-estructurales de carácter paralelo en las cuales se desarrollen procesos de tipo continuo. En este sentido, nos parece especialmente pertinente el análisis de las características de un conjunto de modelos que han sido agrupados bajo la etiqueta de "modelos distribucionales". Como se expondrá en este capítulo, el aspecto más prometedor de este tipo de modelos es su sugestivo "poder emergente": a partir de un sistema arquitectónico relativamente simple, estos modelos parecen ser capaces de aproximarse de forma adecuada a fenómenos muy dispares que forman parte esencial del sistema cognitivo humano.

Los modelos distribucionales vienen siendo propuestos inicialmente, en trabajos recientes, como alternativa a la "metáfora espacial" en el campo de la memoria. Es bien sabido que la investigación sobre la naturaleza de la memoria a largo plazo ha sido siempre un tema central para la psicología cognitiva. Se han propuesto, por citar solamente algunos ejemplos, modelos basados en vectores de componentes (Norman y Rumelhart, 1970), espacios multidimensionales (Shepard, 1962), estructuras en árbol (Mandler, 1967), redes de relaciones semánticas (Quillian, 1968), esquemas, scripts y planes (Bartlett, 1932; Minsky, 1975; Schank y Abelson, 1977; Schank, 1982), series de producciones (Newell, 1973;

Allport, 1980; Anderson, 1983), etc. La mayoría de estos enfoques suelen partir de la base de que la información almacenada en memoria está localizada en algún lugar específico; se han propuesto, a este respecto, comparaciones del sistema de memoria con un archivador, una biblioteca, los registros de memoria de un ordenador digital, etc. Todos los elementos comparativos tienen en común el hecho de basarse en una "metáfora espacial" en virtud de la cual una información específica es almacenada en un locus también específico como, por ejemplo, una rama de un árbol o red asociativa. Cualquier tipo de proceso de recuperación debe basarse, entonces, en la localización del lugar de almacenamiento de la información deseada. En la terminología de las redes semánticas se tratará de inspeccionar los enlaces adecuados; en un sistema multidimensional lo importante será "hallar la región adecuada en el espacio"; en los modelos basados en esquemas será necesario encontrar el esquema apropiado. El principio "localizacionista" se mantiene básicamente invariable, al margen de que se postulen relaciones, de tipo asociativo o de cualquier otra índole, entre los distintos ítems almacenados en el sistema.

En contraste con las aproximaciones anteriores, los modelos distribucionales rechazan explícitamente la "metáfora espacial" y parten de la base de que los ítems almacenados en memoria son más "evocados" o "reconstruidos" que "hallados". La información estaría almacenada, desde este punto de vista, en forma de relaciones entre unidades elementales, de



manera que cada unidad puede participar en la codificación de ítems de información distintos. Como se verá más adelante, este sistema distribucional puede recuperar ítems individuales de forma parecida a cómo un filtro es capaz de extraer componentes individuales de frecuencia a partir de una onda acústica compleja. Además, la estructura de este tipo de modelos es particularmente útil para aprehender la interacción entre ítems; en particular, esta interacción se produce de tal modo que los aspectos comunes entre los distintos ítems se ven reforzados, mientras que los caracteres diferenciales pierden relevancia. Como se indicará algunos párrafos más abajo, esta propiedad permite a estos modelos aproximarse de manera muy eficiente (y con consecuencias teóricas importantes) a los procesos de construcción de conocimiento abstracto.

Como señalan Rumelhart y Norman (1981), este tipo de modelos no se diferencia de construcciones teóricas anteriores solamente en base a sus presuposiciones sobre el almacenamiento de información en memoria, sino en función de otras características particulares. Por una parte, se trata de modelos que toman en gran consideración el criterio de plausibilidad neurológica. En este sentido, se diferencian radicalmente de los modelos computacionales puesto que éstos, como se indicó en otro lugar, solamente introducen restricciones extremadamente laxas sobre el nivel de implementación material. Por otra parte, la posición funcionalista adoptada por los teóricos computacionales niega la relevancia de las

características propias del nivel de implementación material en lo que se refiere a la construcción de modelos computacionales de los procesos cognitivos. Contrariamente, los modelos distribucionales parten de la base de que la arquitectura cerebral introduce límites fundamentales a la forma en que la información puede ser almacenada por el sistema cognitivo humano y, por extensión, sobre el funcionamiento de los procesos cognitivos.

En realidad, este tipo de modelos debe su origen, en parte, a argumentos basados en la fisiología (Anderson y Hinton, 1981). Pueden citarse, como ejemplo ilustre, las reveladoras observaciones de Lashley (1950) en una de sus obras fundamentales ("In Search of the Engram"):

"No es posible demostrar la localización aislada de un trazo de memoria en ningún lugar del sistema nervioso... El recuerdo implica la acción sinérgica o algún tipo de resonancia entre un gran número de neuronas... Dadas las relaciones numéricas implicadas, creo que incluso la reserva de sinapsis individuales para reacciones asociativas especiales es imposible" (pág. 477 y ss.)

Este planteamiento ha servido de base a multitud de trabajos en el ámbito de la fisiología. Dado que no pretendemos realizar una revisión de este campo de trabajo, nos limitaremos a citar, a título de ejemplo, los estudios de Sheperd

(1979), Petsche (1978), Kuffer y Nichols (1976) o Anderson y Hinton (1981), que abogan en favor de la existencia de las propiedades de paralelismo y distribucionalidad en el funcionamiento del neocortex. En una línea semejante se encuentra todo el trabajo de la "escuela pribramiana" en cuanto al carácter holográfico (u "holonómico") del almacenamiento de información en el sistema de memoria. Como es sabido, esta orientación postula que el análisis tipo autómeta del funcionamiento del cerebro resulta insuficiente para dar cuenta de sus propiedades dinámicas, y propugna la construcción de modelos que incorporen propiedades como el carácter distribucional del almacenamiento de información y la consideración del sistema de recuperación más como un conjunto de procesos reconstructivos de la información que como sistemas de búsqueda (Pribram, 1975, 1977; Pribram y Martín Ramírez, 1980).

Algunos argumentos fisiológicos de tipo general juegan un papel importante al explicar el énfasis en el procesamiento en paralelo en oposición al enfoque secuencial propio de otros modelos. Mientras que los modernos ordenadores funcionan a base de unidades de procesamiento que operan en tiempos cercanos a los 50 nanosegundos, el cerebro consta de unidades que trabajan en tiempos del orden de milisegundos. Sin embargo, el cerebro es capaz de ejecutar en periodos de tiempo muy pequeños tareas de procesamiento que pueden exigir muchos minutos (e incluso horas) de tiempo para la unidad central del computador. La consecuencia obvia es que el

cerebro opera simultáneamente con un gran número de unidades elementales de procesamiento.

De hecho, en ocasiones se argumenta que la diferencia entre procesamiento serial y paralelo no es fundamental. Por ejemplo, la teoría de autómatas demuestra que, en principio, una máquina serial puede computar cualquier función asequible para una arquitectura paralela. Pero, como señala Fahlman (1981), el tiempo de computación tiene un efecto tremendo sobre el tipo de computación más efectiva: hay muchos algoritmos que tienen sentido en una máquina paralela, pero que son inviables, en la práctica, en una máquina serial. Una operación de este tipo es, por ejemplo, la intersección de conjuntos, que presenta un coste de tiempo prohibitivo en una máquina serial. Pero el hecho más importante radica en la relación que existe entre el tamaño de los conjuntos intersectados y el tiempo que se necesita para llevar a cabo la tarea. En una máquina serial se produce un crecimiento exponencial del tiempo en función del tamaño de los conjuntos; en algunas máquinas paralelas la relación es distinta e incluso no existe. En términos más generales, y retomando una idea fundamental expuesta en el apartado 2.1, es necesario resaltar que una arquitectura funcional y material de tipo paralelo, como la que postulan los modelos distribucionales, presenta un conjunto de operaciones primitivas que puede ser muy distinto de la constelación de operaciones que realiza una arquitectura de tipo serial (Anderson y Hinton, 1981). Como señalábamos en el punto 2.1,

la esencia del requisito de equivalencia funcional fuerte entre el sistema cognitivo humano y cualquier tipo de modelo que de él se construya radica en la equivalencia entre las operaciones primitivas que ambos sistemas ejecuten (o, en otros términos, en la existencia de restricciones comunes en las arquitecturas funcionales de los dos sistemas). Son estas operaciones primitivas, en tanto en cuanto son definitorias de las propiedades generales del nivel algorítmico (Marr, 1982), las que determinan las características fundamentales del funcionamiento real del sistema y, en consecuencia, las que introducen la posibilidad de contrastación empírica de los modelos que puedan proponerse. Aunque, como acabamos de señalar, una arquitectura serial y otra paralela pueden ejecutar las mismas funciones computacionales, esto no hace sino definir una equivalencia funcional débil entre ambos sistemas y, en particular, no es suficiente para hallar elementos de contrastación empírica entre ellos.

Pensamos, en consecuencia, que cualquier modelo de los procesos cognitivos deben hacer referencia inexcusable a la arquitectura (funcional y material) del sistema, puesto que en caso contrario será muy difícil hallar criterios de evaluación del modelo que vayan más allá de referencias más o menos vagas a la parsimonia de las distintas construcciones teóricas e, incluso, a consideraciones estéticas (Anderson, 1978).

La noción de procesamiento continuo ha sido propuesta (McClelland, 1979) como alternativa a los modelos de



estadios discretos de procesamiento. Sintéticamente, las presuposiciones esenciales de los modelos de estadios discretos, ampliamente mayoritarios en el ámbito de la psicología cognitiva, pueden resumirse en una presuposición principal y una suposición complementaria. La presuposición principal, debida a Donders, es la de que existen fases funcionales sucesivas entre la presentación de un estímulo y la emisión de una respuesta por parte de un sujeto. El término "sucesivas" debe ser tomado en sentido estricto, es decir, indicando que una fase de procesamiento cualquiera no entra en acción hasta que la fase anterior no ha finalizado completamente. En otros términos, la transmisión de información desde una fase de procesamiento a la siguiente se produce de forma "todo o nada", en el sentido de que no existe transmisión de información hasta que la fase antecedente ha finalizado su acción; en ese momento, el output de esa fase pasa a estar disponible para el siguiente estadio de procesamiento. La suposición complementaria indicaría que los componentes del tiempo de reacción son estocásticamente independientes; es decir, la probabilidad de que cierto estadio necesite un tiempo  $t$  para llegar a su finalización es independiente del tiempo que hayan necesitado los demás estadios.

Una segunda suposición complementaria, más fuerte que la anterior, se referiría a la forma de las distribuciones de probabilidad asociadas a cada uno de los estadios. Esta suposición complementaria ha sido analizada y desarrollada por McGill (1963), Kleinrock (1975), Wickens (1982) y otros.

Presentaremos una parte de este análisis, puesto que juega papel esencial tanto para los modelos de estados de procesamiento como para los de procesamiento continuo, ya que ambos modelos necesitan esencialmente derivar una expresión matemática de la distribución de probabilidad para la duración de cualquier operación de procesamiento, al margen de que estas operaciones sean organizadas de forma serial o paralela.

En términos de la teoría de la probabilidad un proceso desarrollado en tiempo continuo es descrito por una distribución de sucesos de Poisson. Un suceso de Poisson se define por el hecho de que la probabilidad de su ocurrencia es proporcional a la longitud del intervalo de tiempo que se tome. Expresado de otra forma:

$$P(\text{un suceso en el intervalo } t, t+\Delta t) = \lambda \Delta t$$

donde  $\lambda$  es una constante que constituirá el parámetro básico del proceso.

Consideremos este proceso como un sistema capaz de generar dos respuestas (o una respuesta y una "no respuesta"): A y A'. El intervalo entre dos respuestas A puede ser considerado como un conjunto de respuestas A'. Si  $p(A) = p$  y  $p(A') = q$ , entonces, si el intervalo se compone de k momentos:  $p(0) = p$ ;  $p(1) = qp$ ;  $p(2) = q^2 p$ , etc. Para cualquier longitud del intervalo:

$P(A/j) = p(j) / \sum_{k=j}^{\infty} p(k)$  ; ya que  $p(k) = q^k p$ , entonces  $P(A/j) = p$ .  
 La expresión  $P(A/j)$  indica la probabilidad condicional de A después de j momentos.

Supongamos que cada respuesta o suceso en el periodo de latencia ocupa un breve intervalo  $\Delta t$ . La probabilidad de que el periodo de latencia finalice,  $p$ , depende de  $\Delta t$ , ya que si  $\Delta t$  disminuye, el número de intervalos  $\Delta t$  en un periodo concreto de latencia aumenta, lo que equivale a decir que  $p$  disminuye. En otros términos:  $p/\Delta t = \lambda$ .

Tomemos la función generatriz de  $p(k)$ :  $M_k(\theta) = p/1 - q e^{\theta}$ . Si  $t = k \Delta t$ , y aplicando el teorema de las funciones generatrices referido al resultado de multiplicar la función original por una constante y sumarle otra constante, resulta la siguiente expresión:

$$M_t(\theta) = p / 1 - q e^{\theta \Delta t}$$

Tomando el límite cuando  $\Delta t$  tiende a cero, y teniendo en cuenta que  $p/\Delta t = \lambda$ , entonces:

$$\lim_{\Delta t \rightarrow 0} M_t(\theta) = \lambda \Delta t / 1 - (1 - \lambda \Delta t + \theta \Delta t)$$

(El término  $\Delta t^2$  es despreciado)

Finalmente :  $\lim_{\Delta t \rightarrow 0} M_t(\theta) = \lambda / \lambda - \theta$

Este resultado es la función generatriz de la distribución exponencial de probabilidad. En efecto, de acuerdo con Laplace, dada una función continua de probabilidad tal que  $f(t)=0$  si  $t < 0$ , podemos definir una función generatriz tal que  $F(s) = \int_0^{\infty} e^{-st} f(t) dt$ .

Ya que la función de densidad de probabilidad de la distribución exponencial es igual a  $\lambda e^{-\lambda t}$ , entonces:

$$M_t(\theta) = \int_0^{\infty} e^{\theta t} f(t) dt = \lambda \int_0^{\infty} e^{-(\lambda - \theta)t} dt = \lambda / \lambda - \theta.$$

En conclusión, la distribución del tiempo esperado de ocurrencia de un suceso de Poisson puede ser descrita adecuadamente por una distribución exponencial de probabilidad.

Obviamente, el suceso esperado en este tipo de análisis no es solamente la respuesta externa final, sino el final de cada una de las fases de procesamiento. En este sentido, la noción de distribución exponencial de la duración de una fase es válida tanto para los modelos de estadios sucesivos de procesamiento como para los de procesamiento continuo, aunque la organización paralela de este último tipo de modelos introducirá algunas particularidades importantes que serán analizadas en su momento.

Puede demostrarse, asimismo, que la distribución de tiempo esperado para un conjunto de  $r$  sucesos de Poisson es descrita correctamente por una distribución gamma con media  $r/\lambda$  y varianza igual a  $r/\lambda^2$  (McGill, 1963). Como veremos más adelante, la distribución gamma juega un papel básico en la ecuación fundamental que define un modelo de procesamiento continuo.

La introducción del concepto de procesamiento continuo y en paralelo es resultado tanto de las dificultades crecientes que los modelos de tipo serial encuentran para dar cuenta de numerosos datos experimentales como de un análisis crítico de sus presuposiciones básicas e incluso de la forma en que abordan el estudio de las distribuciones de tiempo de reacción. Un buen ejemplo de este último tipo de argumentos se encuentra en el análisis propuesto por Ratcliff y Murdock

(1976) y Ratcliff (1979). Estos autores critican el análisis de las distribuciones de TR realizado por Sternberg (1964, 1969, 1975) en base a tres consideraciones fundamentales: la gran magnitud de la varianza asociada con las estimaciones de los momentos de orden alto (lo cual exige un gran número de observaciones), la gran sensibilidad de estas estimaciones a cierto tipo de perturbaciones, y la escasa economía del enfoque, puesto que las estimaciones de los momentos de alto orden proporcionan información sobre una parte de la curva de frecuencias que es de poco interés teórico. A partir de estas y otras consideraciones, Ratcliff y Murdock (1976) argumentan que la convolución de distribuciones normales y exponenciales puede proporcionar una mejor descripción de la distribución de tiempos de reacción. Como es sabido, la convolución nos proporcionaría la distribución de la suma de una variable distribuida normalmente y otra distribuida exponencialmente. En base a este análisis, Hockley y Corballis (1982) hallan que el incremento lineal en el tiempo de reacción medio como función del número de ítems de la lista inicial en una tarea de reconocimiento es resultado de un incremento en la asimetría positiva de la distribución de TR (el parámetro tau en el modelo de convolución), y argumentan que este resultado es problemático para los modelos más habituales de procesamiento serial. Ratcliff (1978) y Hockley (1982, 1984) llegan a conclusiones parecidas utilizando diversas variantes de la tarea de reconocimiento. En particular, Hockley (1984) halla buenos



niveles de ajuste entre las distribuciones empíricas de TR para cuatro tareas distintas y las correspondientes distribuciones teóricas basadas en el modelo de convolución. Los datos son mucho más confusos en cuanto a su compatibilidad con los diversos modelos alternativos propuestos. Así, mientras que una tarea simple de búsqueda visual los resultados se ajustan razonablemente a un modelo serial auto-limitado, en otras tareas los datos pueden ser interpretados tanto desde un modelo serial modificado en el cual la tasa de búsqueda y comparación sería muy variable como, de forma más verosímil, desde un modelo de procesamiento en paralelo.

Meyer, Yantis, Osman y Smith (1984, 1985) realizan un análisis complejo sobre la compatibilidad de los distintos modelos de procesamiento con los datos experimentales. De acuerdo con estos autores, la diferencia entre modelos discretos y continuos se expresaría en términos del número de outputs parciales y transmisiones de información que se producen en el intervalo entre estímulo y respuesta. Obviamente, para un modelo de estadios discretos existirá un número finito de outputs parciales, mientras que en los modelos continuos ese número sería infinito. La propuesta de Meyer et al. (1984) radica, básicamente, en la utilización de una tarea experimental en la cual un primer estímulo proporcionaría un cierto grado de información sobre un segundo estímulo, ante cuya presencia los sujetos experimentales deberían emitir su respuesta. En este tipo de tareas, el primer estímulo puede reducir la incertidumbre en cuanto al tiempo que tardará en

aparecer el segundo (Bertelson, 1967), en cuanto a la localización espacial de este segundo estímulo (Posner, Snyder y Davidson, 1980), o sobre su identidad (Posner y Snyder, 1975). Típicamente, la influencia del estímulo inicial aumenta a medida que crece el intervalo entre este estímulo y el ítem de prueba, hasta llegar a un nivel asintótico (Bertelson, 1967; Neely, 1976; Posner y Snyder, 1975). Además, Meyer et al. (1984, 1985) intentan valorar la influencia del grado de información proporcionado por el estímulo inicial sobre el tiempo de reacción ante el segundo estímulo.

Los distintos modelos de procesamiento permiten derivar predicciones distintas sobre la distribución de tiempos de reacción. La idea esencial es que esa distribución no es sino el resultado global de un conjunto de distribuciones subyacentes y, en consecuencia, es posible, en general, llevar a cabo una descomposición de la distribución original en todas sus distribuciones-base (Everitt y Hand, 1981). La relevancia de este tipo de análisis para la problemática que estamos estudiando es bastante obvia. Los modelos de estadios discretos predecirán que la distribución de tiempos de reacción poseerá un número finito de distribuciones-base, cada una de las cuales corresponderá a un estado de preparación distinto. Existirá, entonces, un estado de preparación total que habrá sido alcanzado si el intervalo entre los dos estímulos es suficiente como para que hayan sido ejecutados todos los estadios de preparación; para intervalos menores el

sujeto habrá alcanzado (dependiendo también del grado de información proporcionado por el primer estímulo) estadios de preparación menos efectivos que existirán solamente en número finito, puesto que el número de estadios de preparación es también finito (y a cada fase serial de preparación alcanzado le corresponderá un cierto estado de preparación). Se supone, lógicamente, que el tiempo de respuesta al segundo estímulo dependerá directamente del estado de preparación alcanzado.

Los modelos de procesamiento continuo permiten derivar predicciones distintas. De acuerdo con este tipo de modelos, en el intervalo entre primer y segundo estímulo se producen cambios graduales en el grado de preparación del sujeto, de forma que la manipulación del intervalo produciría una distribución de tiempos de reacción que no podría ser descompuesta en un número finito de distribuciones-base.

El análisis de Meyer et al. (1985) lleva a estos autores a conclusiones bastante reveladoras. Básicamente, puede afirmarse que los dos tipos de modelos considerados encuentran apoyo en los datos empíricos. Lo fundamental son las condiciones en que se dan ambos tipos de procesamiento: cuando existen solamente dos alternativas de respuesta, los modelos de mejor ajuste son los de tipo serial, con dos o tres estadios de preparación; pero cuando el número de respuestas se incrementa, los modelos que mejor se ajustan a los datos son los de tipo continuo. Dada la simplificación propia de los experimentos psicológicos, no resulta extraño que los

modelos seriales hayan encontrado un buen nivel de soporte empírico. Parece razonable pensar, sin embargo, que experimentos algo más complejos pueden revelar (como de hecho, está ocurriendo en buena medida) la presencia de procesamiento continuo y de procesos al menos parcialmente paralelos. Algo parecido podría decirse en cuanto al efecto de factores como la compatibilidad de respuestas y su complejidad (Meyer et al., 1985). En particular, los datos parecen apoyar la idea de que las respuestas más complejas (por ejemplo, las que exigen un mayor nivel de control muscular) son preparadas mediante un proceso que no admite una descripción en forma de un número finito de fases de procesamiento. Debe señalarse que este efecto parece presentarse incluso en respuestas tan simples como la de pulsar una tecla con el dedo medio que, sin embargo, muscularmente es más compleja que la respuesta estándar con el dedo índice (Marble, 1960). Este tipo de datos sugiere poderosamente la necesidad de revisar la evidencia experimental obtenida en favor de los modelos de estadios de procesamiento, puesto que ésta puede basarse, en buena parte, en el uso de tareas experimentales excesivamente simplificadas.

A pesar de la relativa complejidad y el rigor matemático de los análisis que hemos planteado hasta el momento, pueden persistir dudas en cuanto a su valor probatorio en favor de los modelos de tipo paralelo y con procesamiento continuo, especialmente desde el momento en que utilizan, de modo casi exclusivo, análisis basados en la variable tiempo. Las

variables temporales han sido sometidas a severa crítica en cuanto a su utilidad como indicadores de la duración de eventos de procesamiento subyacentes. Pachella, Smith y Stanovich (1978) sugieren, a partir del planteamiento de una curva idealizada de rapidez-exactitud (figura 4.1), que las dificultades de interpretación de las diferencias cuantitativas entre latencias de respuesta distintas obliga a considerar únicamente sus propiedades ordinales. Una idea básica de este tipo de planteamiento radica en la dificultad de asegurar un rendimiento óptimo del sujeto, en el sentido de que su latencia de respuesta refleje únicamente la duración del procesamiento efectiva y óptimamente dirigido a la tarea planteada. La optimización del rendimiento se verá afectada de forma substancial por el mayor o menor énfasis que presenten las instrucciones en cuanto a los aspectos de rapidez y exactitud. En el caso extremo en que se consiga una ejecución sin errores el problema radica en el hecho de que existe un número infinito de tiempos de reacción compatibles con un rendimiento perfecto (ver figura).

A pesar de la coherencia del planteamiento de Pachella et al. (1978) y a que posiblemente su aceptación tenga bastante sentido en el ámbito, por ejemplo, de la psicología sensorial (Malapeira, 1987), algunas de las conclusiones derivadas por los propios autores a partir de su análisis nos parecen excesivamente radicales. En particular, de la crítica a la utilización aislada del tiempo de reacción como indicador de la duración de supuestos procesos mentales no se



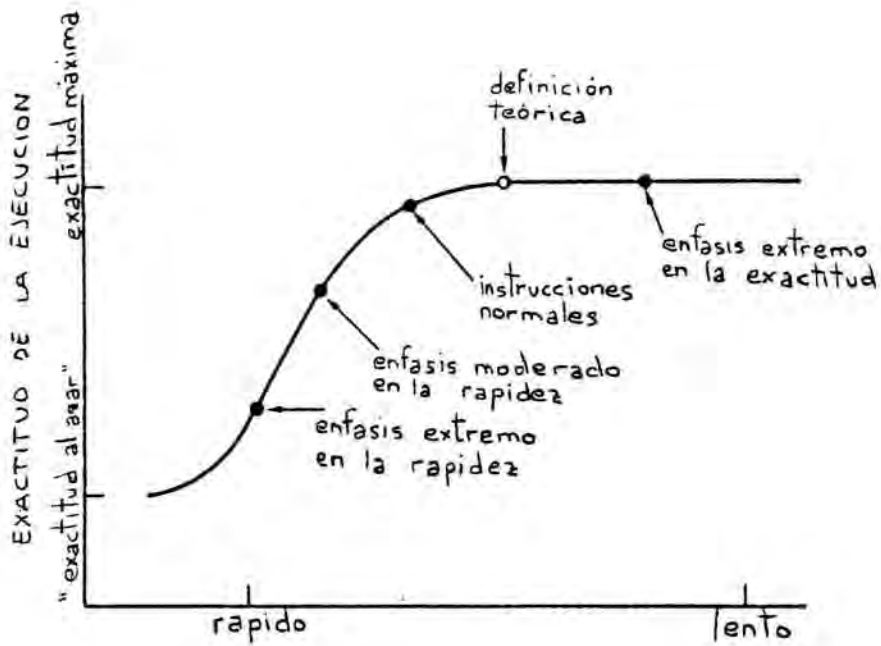


Figura 4.1 Curva idealizada de rapidez-exactitud según Pachella, Smith y Stanovich (1978)

deriva necesariamente la afirmación de que pueda recurrirse únicamente a propiedades de una escala ordinal. Si se plantea, por ejemplo, un modelo a partir del cual puedan derivarse predicciones en cuanto a la relación entre latencia de respuesta y exactitud en la ejecución, entonces parece tener pleno sentido la conservación de las propiedades métricas de la variable tiempo en tanto que escala de intervalos e incluso de razón. El propio análisis de Pachella et al. (1978) y, en particular, su curva idealizada rapidez-exactitud, parten de la existencia de relaciones cuantitativas precisas entre estos dos aspectos. En cualquier caso, sí parece necesario recurrir al análisis de la función rapidez-exactitud como forma de superación de los problemas que plantea el uso aislado de los tiempos de reacción. En este sentido, los modelos de tipo paralelo-contínuo pueden obtener un grado importante de apoyo, puesto que uno de sus planteamientos básicos ha sido precisamente el de derivar predicciones precisas sobre la relación entre latencia de respuesta y exactitud en la ejecución.

Un buen ejemplo de esta idea se encuentra en los trabajos de McClelland (1976, 1979) sobre las propiedades temporales de los procesos mentales. A partir de la crítica de las suposiciones de los modelos de estadios discretos de procesamiento, McClelland introduce la noción fundamental de procesos en cascada, de acuerdo con la cual los distintos procesos implicados en el tratamiento de la información (o al menos la mayoría de ellos) funcionarían de modo paralelo y

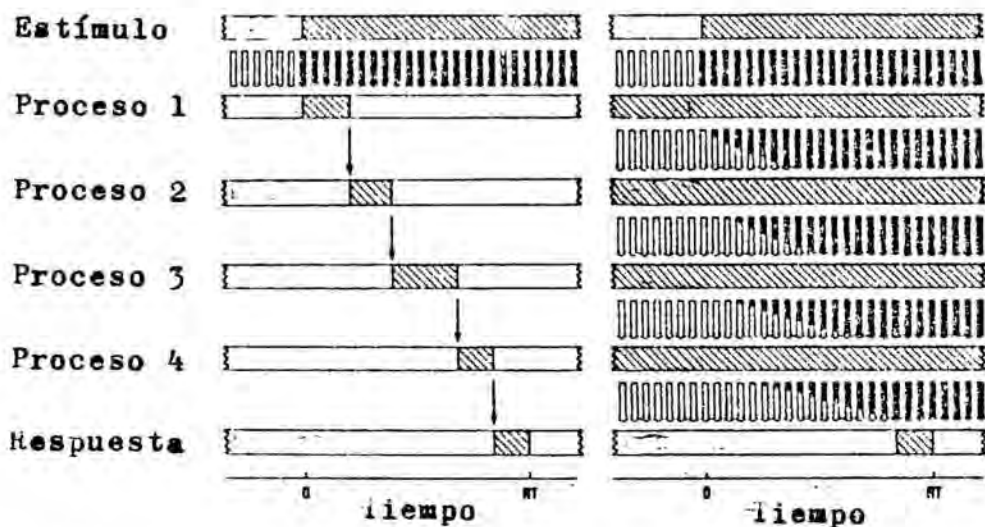
transmitirían de forma continua información sobre el resultado de su actividad. Si consideramos al sistema cognitivo como una estructura jerárquica que incluye diversos niveles, los procesos más complejos irían entrando en funcionamiento a medida que empezaran a recibir información sobre el resultado de actividades de procesamiento situadas en niveles jerárquicos inferiores. La entrada en acción de un elemento concreto de procesamiento dependería, entonces, de la rapidez en el cambio producido en niveles anteriores. En la figura 4.2 se presenta un esquema general del funcionamiento de este tipo de sistema.

En un modelo de procesos en cascada se asume que la información es representada en forma de activación de unidades aisladas o de conjuntos de tales unidades. Estas unidades poseen, de acuerdo con el modelo, el carácter de integradores lineales, es decir, su nivel de activación viene determinado por la suma ponderada de los outputs de las unidades que le suministran información. Por tanto:

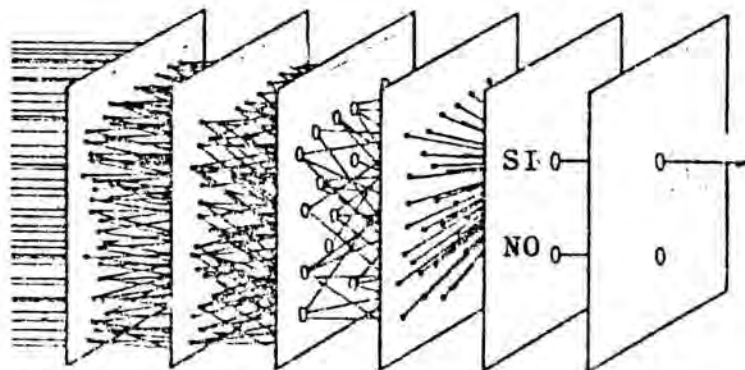
$$A_{nj} = \sum_{j'} W_{j'j} A_{(n-1)j'}$$

donde  $A_{nj}$  representa el nivel de activación de la unidad  $j$  perteneciente al nivel  $n$ , y  $W_{j'j}$  es el peso o influencia de la actividad de  $j'$  sobre el nivel de activación de  $j$ . Resulta claro, ya en primera instancia, que los cambios en el sistema cognitivo (producidos, por ejemplo, por el aprendizaje) tomarán en este contexto la forma de cambios en las conexiones entre unidades, ya sea por aumento o disminución de los pesos, por establecimiento de conexiones nuevas.

Estadios discretos      Modelo en cascada



(a) Sucesos que ocurren en el sistema cognitivo ante la presentación de un estímulo. Las flechas representan la transferencia de información entre procesos.



(b) Selectividad progresiva entre niveles en una tarea de decisión léxica. El nivel más externo corresponde a la detección de luz; el último nivel a la derecha es el de ejecución de respuesta (Si-No\_

Figura 4.2 representación de la acción de un modelo en cascada.

Aunque McClelland no cita esta posibilidad, pensamos que la formación de nuevas conexiones puede ser tratada como un caso extremo de cambio de pesos; en este caso, la conexión entre  $j$  y  $j'$  pasaría de poseer un peso 0 a un peso cualquiera distinto de cero. Este proceso podría poseer entonces un carácter gradual.

En un modelo de sistemas en cascada supondremos también que la tasa de cambio en el nivel de activación de una unidad dependerá de la diferencia entre el input que llegue a ella y su nivel anterior de activación. Esta idea puede expresarse en la siguiente ecuación:

$$\frac{d}{dt}(An_j(t)) = Kn_j (In_j(t) - An_j(t))$$

donde  $K$  es una constante asociada a la unidad  $j$  que representa su "rapidez de reacción" ante la recepción de inputs, mientras que  $In_j$  representa el input total recibido por la unidad.

A partir de las ecuaciones anteriores es posible derivar la expresión fundamental del modelo, es decir, la ecuación de cascada. Para ello sería suficiente con resolver la ecuación de tasa de cambio para el nivel 1, incluir los resultados en la ecuación para el nivel 2 y resolver de nuevo, etc. Sin embargo, es posible obtener una solución general para cualquier nivel arbitrario. El conjunto de las activaciones de los distintos integradores lineales de un cierto nivel  $n$  puede ser expresado como un vector  $A_n = (A_{n1}, A_{n2}, \dots, A_{nN})$ , donde  $N$  es el número de unidades del nivel  $n$ . Igualmente, los pesos asociados a los distintos pares de unidades  $j'j$ ,



(estando  $j'$  en el nivel  $n-1$  y  $j$  en el nivel  $n$ ) pueden ser expresados como una matriz  $M_n$  con  $N'$  filas y  $N$  columnas ( $N'$  es el número de unidades del nivel  $n-1$ ). Entonces, si denominamos  $A_0$  al input que recibe el primer nivel de procesamiento, la activación asintótica en el nivel  $n$  vendría dada por la siguiente expresión:

$$A_n = A_0 \prod_{i=1}^n M_i$$

La ecuación de tasa de cambio tomaría entonces la siguiente forma:

$$\frac{d}{dt} (A_n) = K_n (A_{n-1}(t) M_n - A_n(t))$$

Resulta útil el uso de la transformada de Laplace de la ecuación diferencial que acabamos de exponer. Esencialmente, se trata de obtener la transformada, realizar un conjunto conveniente de manipulaciones algebraicas y obtener la solución final invirtiendo la transformación.

Sea  $f(t)$  una función de  $t$  para  $t \geq 0$ . Entonces la transformada de Laplace de  $f(t)$  viene dada por la expresión:

$$L\{f(t)\} = f(s) = \int_0^{\infty} f(t) e^{-st} dt$$

Un aspecto muy interesante de la transformación de Laplace reside en que la transformada de la derivada de la función original se relaciona de forma algebraicamente bastante simple con la transformada de la propia función:

$$L\left[\frac{d}{dt} f(t)\right] = s\bar{f}(s) - f(0) \quad ,$$

donde  $s$  es una dummy variable y  $f(0)$  es el valor de la función  $f(t)$  para  $t = 0$ .

Entonces, la ecuación de tasa de cambio tomará la siguiente forma:

$$s \bar{A}_n(s) - A_n(0) = K_n(\bar{A}_{n-1}(s)M_n - \bar{A}_n(s))$$

En el caso simplificado de que las activaciones iniciales sean nulas, la expresión anterior se reduce a la siguiente (dado que  $A_n(0)=0$ ):

$$s \bar{A}_n(s) = K_n(\bar{A}_{n-1}(s)M_n - \bar{A}_n(s))$$

Cambiando de miembro algunos términos y dividiendo entre  $(s + K_n)$ , se obtiene:

$$\bar{A}_n(s) = \bar{A}_{n-1}(s) \left[ \frac{K_n}{s+K_n} M_n \right]$$

De la misma forma es posible derivar una expresión para  $A_{n-1}(s)$  en función de  $A_{n-2}(s)$ , y así sucesivamente. Tras la hipotética repetición de este proceso, la ecuación anterior toma la siguiente forma:

$$\bar{A}_n(s) = \bar{A}_0(s) \left[ \prod_{i=1}^n (K_i/s+K_i) \prod_{i=1}^n M_i \right]$$

Realizando una pequeña simplificación, consideraremos que el estímulo a procesar es presentado instantáneamente en el momento  $t = 0$  y permanece en las mismas condiciones hasta que el proceso ha finalizado. Entonces, el valor de  $A_0(t)$  vendrá dado por una expresión del tipo  $u(t)A_0$ , donde la función  $u(t)$  toma un valor 0 cuando  $t < 0$  y un valor de 1 si  $t \geq 0$ . Ya que  $L(u(t)) = 1/s$ , Por tanto, la expresión anterior se convierte en la siguiente ecuación:

$$A_n(s) = (1/s) \prod_{i=1}^n (K_i/s+K_i) \left[ A_0 \prod_{i=1}^n M_i \right]$$

El último paso consistirá en invertir la transformación realizada al principio del proceso. Este problema puede ser resuelto utilizando el método de fracciones parciales de Heaviside. De acuerdo con esta técnica, puede demostrarse que:

$$(1/s) \prod_{i=1}^n K_i/s+K_i = (1/s) - \prod_{l \neq 1}^n (K_l/K_l-K_1)(1/s+K_l) -$$

$$- \dots - \prod_{l \neq i}^m (K_l / (K_l - K_i)) (1/s + K_i) - \dots -$$

$$- \prod_{l=m}^m (K_l / (K_l - K_n)) (1/s + K_n) = (1/s) - \sum_{i=1}^m \prod_{l \neq i}^m (K_l / (K_l - K_i)) (1/s + K_i).$$

La "transformada inversa" de  $1/s$  es igual a 1 (para valores de  $t$  iguales o mayores a 0), y para  $1/(s+K_i)$  el valor de esa transformación es de  $e^{-K_i t}$ . De esta forma:

$$A_n(t) = A_0 \prod M_i \left( 1 - \sum_{i=1}^m \left( \prod_{l \neq i}^m K_l / (K_l - K_i) \right) e^{-K_i t} \right).$$

Para una unidad  $j$  cualquiera perteneciente al nivel  $n$ , la activación asintótica ante la presentación de un estímulo dado  $S$  es igual al elemento  $j$  del vector  $A_0/S \prod_{i=1}^m M_i$ . Entonces, la función de activación para esa unidad viene dada por la siguiente expresión, que constituye la ecuación de cascada que se buscaba:

$$A_{nj}/s^{(4)} = A_{nj}/s \left( 1 - \sum_{i=1}^m \left( \prod_{l \neq i}^m K_l / (K_l - K_i) \right) e^{-K_i t} \right).$$

Puede observarse que la última parte del miembro derecho de la ecuación  $(1 - \dots)$  es idéntica a la forma acumulativa de la función gamma (McGill, 1963). Resulta fácil demostrar que esta parte de la ecuación es igual a 0 cuando  $t=0$  y tiende a un valor asintótico de 1 cuando  $t$  tiende a infinito. Este hecho concuerda perfectamente con la lógica subyacente a las modificaciones de activación de las unidades del modelo.

Retomando la cuestión que comentábamos anteriormente, a partir de la ecuación de cascada pueden derivarse (introduciendo algunas presuposiciones suplementarias) valores para la función que relaciona tiempo de reacción y exactitud en la ejecución. De esta forma, una de las objeciones fundamentales realizadas a la utilización de los tiempos de reacción

como indicadores de procesos psicológicos quedaría, a nuestro entender, seriamente debilitada. En términos generales, la derivación de la curva rapidez-exactitud se basa en las siguientes ideas: Si el sujeto debe seleccionar entre dos respuestas, deberemos suponer que ejecutará la respuesta cuyas correspondientes "unidades representativas" presenten mayor activación. Sin embargo, deben considerarse al menos dos fuentes de variabilidad en esa activación: por una parte, la "activación de base" de esas unidades, que se referiría a la activación producida por factores ajenos al input que recibe cada unidad (en el caso de las unidades de respuesta puede pensarse, por ejemplo, en la existencia de sesgos previos en favor o en contra de cada una de las respuestas posibles); por otro lado, una segunda fuente de variabilidad podría encontrarse en los propios estímulos (por ejemplo, diferentes ítems "nuevos" y "viejos" en una tarea de reconocimiento).

Dada la existencia de fuentes de variabilidad (que no deben agotarse necesariamente en las dos ya expuestas), parece pertinente un análisis basado en la teoría de detección de señales. Consideremos una típica tarea que admita únicamente dos respuestas (por ejemplo, si-no). La idea sería entonces que existe un cierto criterio de forma que siempre que la activación de las unidades representativas de la respuesta "si" supere dicho criterio el sujeto responderá afirmativamente, mientras que en caso contrario la respuesta será "no". Obviamente, esta idea es generalizable a tareas de

indole muy distinta: por ejemplo, una tarea de reconocimiento puede ser concebida como una serie de decisiones si-no referidas a cada uno de los ítems de la lista de reconocimiento. Dadas estas consideraciones, es factible usar la teoría de detección de señales para derivar una fórmula para la exactitud de las respuestas iniciadas en cualquier momento  $t$  que se desee considerar:

$$d'(t) = (\hat{A}s.s(t) - \hat{A}s.n(t)) / \sigma_T(t) \quad ,$$

donde  $d'$  representa una medida de la exactitud de las respuestas emitidas en el momento  $t$ ;  $A.s.s(t)$  sería el valor medio de la activación de la unidad (o unidades) de respuesta "si" bajo un estímulo que exige una respuesta afirmativa;  $A.s.n(t)$  es la activación media de esa misma unidad ante un estímulo que exija respuesta negativa, y  $\sigma_T$  se derivaría de las varianzas de las activaciones bajo las dos circunstancias mencionadas, de acuerdo con la siguiente expresión:

$$\sigma_T(t) = \sqrt{\frac{\sigma_{s/s}(t)^2}{2} + \frac{\sigma_{s/n}(t)^2}{2}}$$

McClelland (1979) demuestra que este tipo de planteamiento permite una interpretación verosímil de un buen número de datos experimentales y hace posible también derivar predicciones en cuanto a la influencia de ciertas manipulaciones experimentales. Por otra parte, el modelo puede ser desarrollado en diversas direcciones; por citar un ejemplo, sería posible, en principio, postular la existencia de un criterio de respuesta fijo (superficialmente similar al criterio "random-walk" adoptado por algunos investigadores; Ratcliff,



1978), o bien de un criterio variable de un ensayo a otro (Grice, 1968).

Señalemos finalmente que el modelo en cascada es compatible con los datos obtenidos a partir del método de factores aditivos de Sternberg; en particular, a partir de un modelo de procesos en cascada puede predecirse que los factores que supuestamente afectan a dos procesos distintos producirán efectos aditivos en los tiempos de reacción, mientras que los factores que afecten a un único proceso tenderán a producir efectos interactivos. A partir de esta consideración y de todas las anteriormente realizadas, pensamos que los modelos distribucionales interactivos pueden constituir una alternativa fundada y poderosa a los modelos de base serial y discreta en el marco de la psicología cognitiva.

Antes de proseguir con la exposición detallada de los aspectos de los modelos distribucionales que pueden ser más relevantes a nuestro trabajo, es necesario referirse a una discusión fundamental respecto al nivel que este tipo de modelos ocupa en el marco de la teoría psicológica. En particular, Broadbent (1985) sostiene la idea de que los modelos distribucionales pertenecen a lo que Marr (1982) denominaría nivel de implementación física; en otros términos, circunscribe el alcance de este tipo de modelos al ámbito de la fisiología. Si se aceptara este argumento, resultaría problemático proponer a los modelos distribucionales como alternativa a ningún modelo psicológico y, en el mejor de los casos, su papel sería el de introducir algunas restricciones

respecto a la verosimilitud de la implementación material de algunos de esos modelos.

McClelland y Rumelhart (1985) afirman explícitamente que los modelos distribucionales son modelos psicológicos y que, como tales, pueden ser comparados con otros modelos del mismo nivel. Por ejemplo, puede establecerse una comparación entre un modelo distribucional y la concepción prototípica de los procesos de categorización y conceptualización. Como se verá en el siguiente apartado, en numerosas situaciones las predicciones que pueden realizarse a partir de ambos modelos son claramente divergentes. En particular, los modelos distribucionales son capaces de predecir las condiciones bajo las cuales la categorización es regulada por un prototipo o bien sigue una dinámica basada en ejemplares. Del mismo modo, un modelo distribucional puede ser comparado perfectamente con la teoría de los "logogenes" de Morton (1969, 1979, 1981) (con la interesante particularidad de que el modelo distribucional puede evitar la proliferación de "logogenes" para distintas variaciones de las condiciones de presentación), o con los llamados "modelos enumerativos" (Jacoby, 1983a, 1983b), sobre los cuales poseen ventaja en cuanto, por ejemplo, a la interpretación de la influencia del conocimiento anterior del sujeto sobre el almacenamiento de nueva información. En todos estos casos, las ideas fundamentales de que las operaciones de procesamiento que ejecuta el sistema son controladas por las conexiones entre un conjunto de unidades elementales de procesamiento y de que la

representación toma la forma de configuraciones de activación sobre esas unidades parecen ser lo bastante poderosas y significativas como para derivar predicciones empíricas que, para muchas situaciones experimentales, divergen de las que pueden obtenerse a partir de otros modelos. A nuestro entender, esta es una importante razón para defender el carácter psicológicamente significativo de los modelos distribucionales.

En su propio análisis, Broadbent (1985) recurre a la noción de niveles de la teoría del procesamiento de información propuesta por Marr (1982), aunque utilizándola de un modo un tanto especial. En particular, Broadbent tiene en cuenta únicamente la existencia de dos niveles de la teoría, a saber, el nivel de teoría de la computación y el de implementación material. Naturalmente, desde esta perspectiva la objeción básica respecto a los modelos distribucionales radicará en su equivalencia computacional con los modelos locales a los cuales pretende substituir. En particular, Broadbent recurre a un argumento que hemos mencionado ya en este trabajo: los sistemas distribucionales no son capaces de llevar a cabo una computación que no sea, a su vez, realizable por un sistema de almacenamiento "localizacionista" (Broadbent, 1985, pág. 190). A este respecto debemos señalar, una vez más, que la equivalencia computacional es débil, es decir, no permite la asimilación de un modelo a otro. En su respuesta a estas consideraciones, Rumelhart y McClelland (1985) insisten en esta misma idea al señalar que

algoritmos de ejecución distintos pueden ser computacionalmente equivalentes y, sin embargo, presentar grandes diferencias a nivel empírico. El recurso a la tesis de Turing puede ser suficiente para establecer la equivalencia a nivel computacional, pero es de poca utilidad para la psicología a causa de que ésta centra su interés en formas de equivalencia fuerte entre procesos psicológicos y los modelos que de ellos podamos construir.

Rumelhart y McClelland (1985) inciden en la insuficiencia del análisis de Broadbent y centran gran parte de sus críticas en la omisión, por parte de éste, del nivel algorítmico propuesto por Marr. En realidad, el nivel algorítmico sería el más genuinamente psicológico, puesto que se plantea dar respuesta a las cuestiones relacionadas con la forma de representación de la información y con los procedimientos utilizados para su transformación. A nuestro entender, resulta bastante claro que los modelos distribucionales realizan propuestas alternativas y altamente específicas para los dos aspectos esenciales del nivel algorítmico: por una parte, la información es representada, de acuerdo con estos modelos, en forma de patrones de activación que admiten propiedades tanto de tipo todo o nada como de tipo continuo (y que, por tanto, son compatibles con representaciones densas); por otro lado, la transformación de esas representaciones consiste en cambios en esos patrones de activación en forma de modificaciones en la "fuerza" de las conexiones entre unidades de procesamiento (un concepto muy semejante al de fuerza

asociativa, un elemento fundamental de muchos modelos psicológicos).

Es cierto que los modelos distribucionales han intentado siempre atenerse a criterios de verosimilitud fisiológica, y es verdad también que su relación con el nivel fisiológico es mucho más cercana que la que puede encontrarse, por ejemplo, en los modelos propuestos por muchos teóricos computacionales. Sin embargo, esta mayor cercanía al nivel de implementación física no equivale a la reducción de los modelos distribucionales a modelos fisiológicos, desde el momento en que no se define la relación entre las unidades de procesamiento propias del enfoque distribucional y las unidades funcionales del sistema nervioso. En otros términos, una unidad elemental de procesamiento podría corresponder a una neurona individual, o a un grupo de neuronas que funcionarán conjuntamente (podría, incluso, proponerse la improbable idea de que el funcionamiento de una unidad elemental de procesamiento fuera el resultado de un patrón de activación "distribuido" a lo largo del sistema nervioso). Además, las conexiones establecidas en un modelo distribucional del tipo que estamos considerando no deben ser necesariamente de la misma naturaleza que las establecidas entre las neuronas. Por ejemplo, en un modelo distribucional podría postularse que cada una de las unidades tiene asociado un "campo de fuerzas" que influye sobre un conjunto de unidades próximas a través de procesos que presenten analogías con las propiedades de un campo físico; como en cualquier otro modelo no



fisiológico, no será imprescindible hacer referencia a la forma exacta en que ese tipo de procesos estén implementados en el sistema nervioso (lo cual no contradice, por supuesto, la afirmación de que el nivel de implementación material pueda, en algún momento, introducir restricciones decisivas para el modelo no fisiológico). En general, nuestra opinión es que mientras no se defina una aplicación entre las unidades de funcionamiento de un modelo distribucional y las del sistema nervioso no es posible definir la forma que toma la relación entre los dos niveles. Lo que ocurre es que, al menos en principio, la relación entre una unidad elemental de procesamiento y el substrato fisiológico debe ser bastante más simple que la que pueda establecerse entre un estado mental complejo y ese mismo substrato, pero esto no significa que en el primer caso deba existir necesariamente reducción a un nivel fisiológico.

#### 4.2 El poder emergente de los modelos distribucionales

Como se ha indicado en varias ocasiones a lo largo del texto, una de las razones que abogan en favor de la consideración y análisis de los modelos de tipo distribucional en el marco de la psicología cognitiva es su sorprendente capacidad para aprehender, a partir de una "estructura arquitectónica" relativamente simple, fenómenos que presentan un alto grado de complejidad. Como se señalaba en el apartado anterior, la aproximación basada en modelos distribucionales ha

demostrado poseer ciertas ventajas sobre modelos sólidamente establecidos como, por ejemplo, la concepción prototípica de la formación de conceptos. A lo largo de este apartado intentaremos mostrar algunos ejemplos relevantes del "poder emergente" de los modelos distribucionales, especialmente en aquellos aspectos que sean de mayor interés para el desarrollo del presente trabajo, intentando al mismo tiempo establecer con mayor claridad algunas de las presuposiciones fundamentales que subyacen a este tipo de modelos y valorar sus implicaciones para el planteamiento que estamos realizando.

Consideraremos en primer lugar algunas peculiaridades de los modelos distribucionales en su calidad de modelos de la memoria humana. Debe señalarse, a este respecto, la existencia de dos tipos básicos de modelos de base distribucional, cuya diferencia radica esencialmente en el tipo de instrumento matemático utilizado en su formulación. Por un lado, Murdock (1982, 1983) y Eich (1982) han propuesto un modelo para los procesos de recuerdo y reconocimiento basado en el concepto de convolución de vectores; de otra parte, Pike (1984), a partir de los trabajos de Anderson, Silverstein, Ritz y Jones (1977) y de Kohonen (1977), propone una alternativa basada en un modelo matricial que, de acuerdo con su autor, presenta menores dificultades matemáticas y es capaz de dar cuenta de mayor número de fenómenos con mayor economía en los sistemas de almacenamiento de la información.

Los modelos "convolucionales" han sido utilizados en el análisis de tareas de recuerdo libre (Metcalf y Murdock, 1981), de ciertas propiedades asociativas de la información (Eich, 1982, Murdock, 1982), de efectos seriales (Murdock, 1983) y, en general, para el estudio de diversos fenómenos relacionados con la codificación, almacenamiento y recuperación de la información (McDowd y Murdock, 1986). La versión más evolucionada de este tipo de aproximación es el modelo TODAM (Theory of Distributed Associative Memory) propuesto por Murdock (1982, 1983).

En el modelo TODAM los ítems son representados como "vectores aleatorios", es decir, vectores cuyos elementos son variables aleatorias, cada una de las cuales corresponde a un determinado atributo de la configuración estimular. Se asume que estas variables están distribuidas normalmente con una media de 0 y una varianza igual a  $P/N$ , siendo  $P$  la potencia del vector y  $N$  el número de elementos que contiene. El mecanismo asociativo fundamental consistirá, entonces, en la convolución de los vectores correspondientes a estímulos particulares (Eich, 1982). La idea es que dos vectores separados se combinarán en un único vector compuesto de dimensión  $2N - 1$ , siendo  $N$  la dimensión de los vectores originales. En este esquema, las características individuales de cada uno de los ítems iniciales pierden su carácter aislado (es decir, no estarían almacenadas en una localización específica), y quedan de alguna forma "distorsionadas" al formar parte de un "conglomerado de información" que contiene las

características de los dos ítems originales. En conjunto, se trata de una idea bastante parecida a las que se han propuesto a partir de la tecnología holográfica (Pribram, 1975; Pribram y Martín Ramirez, 1980). El proceso de convolución no finaliza en este punto, sino que posee un carácter general, de modo que sucesivos pares de ítems (combinados ya por convolución) pueden ser combinados en un nuevo conglomerado de información, y así sucesivamente. De esta forma, el resultado final será un vector único  $M$  que contendrá tanto la información referida a cada uno de los ítems originales como la información relativa a las relaciones asociativas entre esos ítems.

El proceso de recuperación de información toma distinta forma según el tipo de tarea propuesta. En el caso de la tarea experimental ideada por Sternberg, por ejemplo, se asumirá que el ítem de prueba será comparado con el vector de memoria correspondiente a la lista original mediante el producto de puntos. El resultado de este producto es un escalar que conducirá a una respuesta positiva o negativa en función de si se encuentra por encima o por debajo de un cierto umbral. En una tarea que implique reconocimiento de orden serial se asumirá que se forman vectores de memoria separados para la lista original y para la lista de prueba. Ambos vectores son comparados entonces mediante el producto de puntos; la toma de decisión se realiza de la misma forma que en la tarea anterior. Es interesante resaltar que este tipo de procedimiento es capaz de funcionar adecuadamente para

diversas variantes de la tarea de reconocimiento de orden serial (por ejemplo, en el caso de que el número de ítems de la lista original y de la lista de prueba no sea idéntico). Esto lleva a Murdock (1983) a señalar que nos encontramos ante un sistema de reconocimiento de propósito general capaz de discriminar entre ítems en tareas de distintas características.

En el caso de tareas de recuerdo con conservación del orden serial, la operación básica de recuperación es la correlación: el primer ítem es correlacionado con el vector de memoria, y el resultado es una aproximación al segundo ítem (y así sucesivamente). En este caso el resultado no es un escalar sino un vector que reproduce (con cierto grado de aproximación, dependiendo del valor de ciertos parámetros del modelo) el conjunto de ítems presentado originalmente.

Por su parte, el modelo matricial de Pike (1984) es semejante al modelo convolutivo en el sentido de que lo que se almacena en memoria es la suma de pares de ítems asociados, pero difiere del modelo anterior en algunos aspectos importantes. A diferencia del análisis propio del modelo convolutivo, basado en la formación de vectores que representan agregaciones de información, en el modelo matricial cada componente asociativo consiste en una matriz de elementos formada a partir de la matriz producto de dos ítems originales. Es decir, para dos ítems  $a$  y  $b$ , cada uno de ellos con  $N$  elementos (atributos), la asociación entre ellos se representará como una matriz  $ab'$  de orden  $N \times N$ . El



componente mnémico global  $M$  vendrá dado entonces por la suma de las matrices asociativas construidas para cada uno de los pares de ítems asociados.

De acuerdo con Pike (1984), la ventaja fundamental de este planteamiento radica en su mayor simplicidad matemática puesto que, por ejemplo, el cálculo del valor escalar que debe determinar la respuesta del sujeto es más simple que en el modelo convolutivo. La propiedad del formalismo matricial que explica su simplicidad analítica radica en que los procesos de recuerdo asociativo o de reconocimiento pueden ser representados como un conjunto de operaciones elemento a elemento establecidas entre dos matrices asociativas. Por otra parte, Anderson et al. (1977) y Hinton y Anderson (1981) sugieren que, en general, un modelo basado en matrices asociativas posee una interpretación neurológica plausible en términos de una matriz de "fuerzas sinápticas". En general, los modelos convolutivo y matricial han sido comparados también en términos de la generación de ruido en el proceso de recuperación, espacio de almacenamiento requerido, efectos de orden, carácter no direccional (modelo convolutivo) o multidireccional (modelo matricial) de las asociaciones, etc. (Pike, 1984; Murdock, 1985; McDowd y Murdock, 1986). En cualquier caso, lo importante aquí es señalar que los modelos de base distribucional vienen mostrando repetidamente su capacidad para dar cuenta de un buen número de fenómenos asociados con el sistema de almacenamiento y recuperación de información.

En un ámbito bastante distinto, Dell (1980, 1984, 1985) presenta un análisis de la aplicación de modelos en paralelo a la producción del lenguaje. La aportación más interesante de estos trabajos es, a nuestro entender, su análisis del papel que juega el feedback positivo entre niveles. Puede citarse, por ejemplo, el que se establece desde el nivel de "palabras" sobre el nivel de "letras": las conexiones excitatorias establecidas desde el nivel de palabras al de letras permiten que el nivel superior afecte el funcionamiento del inferior. Supongamos, por ejemplo, que en el nivel de análisis de letras se encuentran patrones de activación de igual intensidad que corresponden a G1 (indicaría que la primera letra de la palabra es una G), C1, A2, B3 y O4. Obviamente, G1 y C1 son incompatibles; el resultado del análisis a nivel de letras puede ser influido por el feedback procedente del nivel superior, según el cual CABO es una palabra y "GABO" no lo es. El resultado de esta influencia será un aumento en la activación de C1, de forma que el nivel de activación de los dos patrones incompatibles se hará desigual. Este tipo de proceso es lo que Feldman y Ballard (1982) denominan "coalición estable" entre niveles. Naturalmente, pueden citarse otros ejemplos de influencia de niveles superiores sobre otros de orden inferior. Por ejemplo, el proceso de percepción de letras puede sufrir restricciones desde un nivel léxico, lo cual será especialmente útil cuando el estímulo presente un alto nivel de ruido; en otros casos, un proceso de "filtraje" puede evitar la

codificación de cadenas de sonidos no consistentes con información de nivel superior (Dell, 1985).

Quizá el aspecto más interesante de este tipo de trabajos es el hecho de que proporcionan un principio general para modelar la influencia de niveles "altos" sobre niveles inferiores. La inclusión de este tipo de relación permite aumentar considerablemente el poder de los modelos basados en conexiones entre unidades dispuestas jerárquicamente. Por ejemplo, el modelo de activación interactiva propuesto por McClelland y Rumelhart (1981) y Rumelhart y McClelland (1982) permite una aproximación excelente a fenómenos como la ventaja perceptual para letras que pertenecen a palabras sobre letras incluídas en contextos no relacionados, y a otros hechos relacionados con el contexto léxico.

En cuanto a la producción de lenguaje, parece que el feedback positivo juega dos funciones principales: en primer lugar, permite una "edición pre-articulatoria" del patrón de activación teniendo en cuenta las restricciones procedentes de niveles superiores (como es el caso de las impuestas desde un "nivel morfémico" sobre el nivel fonológico). Este tipo de "mecanismos editoriales" permite evitar errores asociados con el carácter inherentemente "ruidoso" de la producción y comprensión del lenguaje (un carácter que, por cierto, es muy bien captado por los modelos que estamos considerando); estos errores pueden referirse, por ejemplo, a los procesos de recuperación (activación de un número excesivo de unidades) o a la variabilidad del input. La segunda

función del feedback positivo sería la de relacionar la selección de palabras con la selección de estructuras sintácticas. Ciertas estructuras sintácticas exigen la selección de ciertos ítems léxicos (por ejemplo, objetos que reciben acción - verbos transitivos). Bock (1982) ha presentado un modelo en el que la selección de palabras y de estructuras sintácticas se produce en paralelo y es influenciada por la existencia de conexiones excitatorias entre las unidades implicadas en ambas funciones.

Es necesario resaltar el poder predictivo del modelo de conexiones jerárquicas propuesto por Dell (1980, 1985). Es posible derivar, a partir de este modelo, predicciones sobre la presencia de ciertos sesgos léxicos, cuya existencia y estabilidad ha sido demostrada por Berg (1983) y Dell y Reich (1981). El modelo puede dar cuenta sin problemas de fenómenos como el efecto de repetición de fonemas sobre el error de "intercambio fonémico" (MacKay, 1970), efectos de facilitación semántica (Meyer y Schvaneveldt, 1971), etc. En general, este modelo interactivo parece ajustarse mejor a los datos experimentales que los modelos de nivel único propuestos para el efecto de fonema repetido (Wickelgren, 1969) o para los sesgos léxicos (Adams, 1979).

Como se indicaba en el apartado 2.4, existe todo un conjunto de fenómenos relacionados con el procesamiento del lenguaje que sugieren la necesidad de acudir a modelos altamente interactivos (que exigen, claramente, la presencia de un importante grado de "paralelismo" en las operaciones de

procesamiento). Citábamos entonces hechos como la ambigüedad del lenguaje, la presencia de ciertos errores de comprensión, la capacidad de interpretar textos no gramaticales, etc. Señalábamos también la relevancia que, a nuestro juicio, poseía la distinción entre sistemas de propagación de activación digitales y analógicos. Las dos formas de propagación presentan algunos problemas comunes, aunque las posibilidades de solución de algunas de estas dificultades parecen ser bastante distintas. Estos problemas parecen referirse, ante todo, al control de la propagación de la activación. En el caso del sistema digital, por ejemplo, pueden ser explorados, a lo largo de la red, muchos "caminos" irrelevantes o de poco interés. En el caso del sistema analógico, el principal peligro radica en lo que podríamos denominar "muerte entrópica", es decir, en la activación uniforme del conjunto de la red.

A este respecto, Waltz y Pollack (1985) proponen un modelo altamente interactivo que incorpora un proceso de inhibición lateral destinado a coordinar las decisiones distribuidas, siempre con el objetivo de evitar la uniformización de la activación de toda la red. La idea de inhibición lateral ha sido tomada, entre otras fuentes, de los estudios de percepción visual y, esencialmente, consiste en la existencia de conexiones inhibitorias destinadas a evitar la ejecución simultánea de procesos opuestos o incompatibles. La idea es que estas conexiones inhibitorias se establecerían entre pares de unidades (o conjuntos de unidades) cuyo



funcionamiento simultáneo esté contraindicado; de esta forma, el aumento en el nivel de activación de uno de los miembros del par conduciría a una disminución en el nivel de actividad del otro miembro del par. Este proceso de "relajación" conduciría, por ejemplo, a la construcción de una interpretación única de cualquier frase que sea introducida en el sistema.

Pensamos que los enlaces inhibitorios propuestos por Waltz y Pollack (1985) juegan un papel complementario al de los enlaces excitatorios propuestos en el modelo de Dell (1985). La utilización combinada de ambos tipos de conexión permitiría que el sistema cognitivo, cuyo estado inicial frente a un input cualquiera podría ser descrito como ambiguo e inestable, lograra generar patrones de activación estables que corresponderían a la interpretación de ese input. En este sentido, suscribimos plenamente la opinión de Winograd (1981) cuando afirma que la presentación de un ítem puede ser considerada como "una perturbación en un sistema cognitivo en funcionamiento que está intentando dar sentido a las cosas" (pág. 245). El problema será entonces el de ajustar el funcionamiento global del sistema a fin de incorporar ese nuevo ítem al conjunto de información que maneja ese sistema en un momento dado.

Hinton (1981) analiza la posibilidad de construir redes semánticas que funcionen mediante procesos en paralelo. En su formulación, un nodo de la red no es identificado con una localización concreta sino con un patrón de activación que

implica el funcionamiento de un conjunto amplio de unidades. En este sentido, la propuesta de Hinton se diferencia de los modelos de "implantación directa" propuestos, por ejemplo, por Collins y Loftus (1975), Fahlman (1981) o Feldman (1981). De acuerdo con Hinton, el formalismo de las redes semánticas puede ser considerado como una descripción poco aproximada de las interacciones entre nodos. En particular, añadimos nosotros, si las "etiquetas" que enlazan distintos nodos son substituídas, como postula Hinton, por conexiones ponderadas entre unidades, y si esas conexiones poseen niveles de acción continuos, entonces la propuesta de Hinton implica, como lo hacen en general este tipo de modelos, una forma de funcionamiento radicalmente distinta, aunque "sólo" sea porque esa clase de conexiones no trabaja en base a propiedades formales de una cadena de símbolos (en este caso, la "etiqueta" que conecta dos nodos) sino en función de una dinámica continua. Naturalmente, esto no descarta la posibilidad de acudir a otro tipo de formalización (por ejemplo, de base matemática) que permita aprehender en mayor medida las propiedades de este tipo de sistemas. Este tipo de formulaciones parece aproximarse mucho a lo que Carello et al. (1982) denominan "proceso dinámico de despliegue" del significado (ver, a este respecto, el apartado 2.2).

Los intentos de representar redes semánticas en arquitecturas de tipo paralelo inciden en algunas cuestiones de importancia. En primer lugar, introducen la demostración básica de que no es posible obtener la propiedad de sensibilidad al

contexto simplemente combinando los efectos de dos patrones de activación. Supongamos que se trabaja con los siguientes "triples" (expresiones simplificadas para evitar complicaciones de procesamiento), cuyo significado es inmediatamente intuitivo:

UNO	IGUAL	UNO
UNO	DIFERENTE	DOS
DOS	DIFERENTE	UNO
DOS	IGUAL	DOS

¿Cómo puede el sistema, en presencia de los dos primeros elementos del triple, activar el patrón correspondiente al tercer elemento?. Supongamos que este tercer elemento corresponde al funcionamiento de cierto conjunto de unidades elementales, y que existe un cierto umbral  $U$  por encima del cual la respuesta de estas unidades es "uno"; si el nivel de activación no llega al umbral, la respuesta de estas unidades corresponde a "dos". Sean  $Q_1$ ,  $Q_2$ ,  $Q_i$  y  $Q_d$  los "pesos" que modulan la influencia de los patrones de activación de "uno", "dos", "igual" y "diferente", respectivamente, sobre las unidades responsables del tercer elemento del triple. Entonces, para que el sistema funcione correctamente será necesario que:

$$Q_1 + Q_i > U$$

$$Q_1 + Q_d \leq U$$

$$Q_2 + Q_i \leq U$$

$$Q_2 + Q_d > U$$

Combinando la primera y la cuarta expresión, resulta:

$$Q_1 + Q_2 + Q_i + Q_d > 2U$$

Pero, combinando las expresiones segunda y tercera, el resultado es:

$$Q_1 + Q_2 + Q_i + Q_d \leq 2U$$

Obviamente, no existe ningún grupo de números que cumpla las dos expresiones obtenidas. En otras palabras, es imposible asignar un valor a los pesos de las conexiones entre las unidades que forman el sistema. Hinton (1979, 1981) muestra cómo la acción de unidades de otros niveles de orden superior, que responden a los patrones de activación procedentes de los elementos del "triple", permiten solventar la dificultad. En particular, Hinton introduce en su modelo un teorema de convergencia, ideado originalmente para el funcionamiento de unidades TLU (Threshold Logic Unit; Rosenblatt, 1961; Nilsson, 1965; Minsky y Papert, 1969), que permite ajustar los pesos asociados a ciertas conexiones para lograr que un determinado input dé lugar a un output concreto. Este tipo de argumentos y de desarrollos técnicos, que pueden ser generalizados a muchos otros casos, apoyan considerablemente las presuposiciones sobre la necesidad de un funcionamiento fuertemente interactivo de este tipo de sistemas.

Otro aspecto muy importante de este tipo de planteamientos se refiere a la forma en que abordan la relación entre "señales" y "tipos", es decir, el problema de la categorización y la generalización. Las redes semánticas "clásicas", basadas en asociaciones etiquetadas, solucionaban este problema remitiéndose a un nodo de un nivel de abstracción adecuado

para resolver la cuestión que estuviera planteada en cada momento. Este tipo de solución jerárquica era obviamente preferible a cualquier otra que implicara la repetición de los caracteres asociados para cada uno de los ejemplos de determinada categoría (Quillian, 1968; Collins y Quillian, 1969). Sin embargo, el planteamiento distribucional-interactivo sugiere sus propias alternativas a esta problemática. Como veremos algunos párrafos más adelante, estas alternativas se basan, generalmente, en la suposición de que la abstracción es el resultado de la superposición de los aspectos comunes de los patrones de activación correspondientes a los ejemplos particulares de cierta categoría.

Los trabajos realizados en torno a las "redes semánticas en paralelo" inciden también en la importante distinción entre el contenido directo de un concepto (el conjunto de características que lo definen) y su contenido asociativo (su relación con otros conceptos). La idea es en buena parte semejante a la distinción propuesta por Mandler (1980) entre estructura integrativa y estructura elaborativa. En esencia, puede afirmarse que un ítem posee tanto una estructura interna como relaciones externas con otros ítems. Lo fundamental es señalar que, desde una visión "reconstructiva" de la memoria, la asociación entre ítems es "recuperada" a través de un proceso que implica la creación de una "resonancia" que depende del contenido directo de los ítems asociados. Es decir, ante la presentación de uno de los elementos de un



par asociado, la recuperación del otro será el resultado de "evocar" el patrón de activación común a ambos ítems, que a su vez toma su forma específica a partir de las propiedades de los elementos originales. Como veremos más adelante, esta "resonancia" puede ser interpretada en términos del establecimiento de un régimen de actividad común a los dos ítems originales. Los modelos que estamos analizando plantean la relación entre los aspectos integrativos (contenido directo) y elaborativos (contenido asociativo) en base al mismo tipo de principios: los conceptos son patrones de actividad y las asociaciones consisten en interacciones entre esos patrones que dan lugar, a su vez, a patrones de activación nuevos o, en los términos que utilizan algunos modelos distribucionales de la memoria, a convoluciones entre vectores que representan conceptos o ítems originalmente no relacionados.

Los modelos de redes semánticas implantados en arquitecturas paralelas plantean, junto a los aspectos mencionados hasta el momento, dos cuestiones de ámbito muy general. En primer lugar, y como ya hemos señalado, la problemática general asociada con los procesos de abstracción, absolutamente fundamentales en la dinámica de los procesos cognitivos; como segundo tema, la cuestión más general de los aspectos adaptativos de la conducta y, en particular, la problemática relacionada con los presumibles cambios que en el sistema cognitivo produce el aprendizaje.

En cuanto al tema de la abstracción, debemos insistir nuevamente en que su planteamiento general, en el marco de los modelos distribucionales-interactivos, consiste en introducir en el modelo procesos destinados a favorecer la activación de los aspectos comunes a distintos elementos y a neutralizar (aunque, como veremos, no eliminar) sus aspectos diferenciales. Esta propiedad es común a los distintos modelos de categorización propuestos en el marco que estamos considerando (Anderson y Mozer, 1981; Knapp y Anderson, 1984; McClelland y Rumelhart, 1985). La idea básica es que las propiedades funcionales equivalentes de las representaciones abstractas -prototipos, "logogenes" e incluso reglas- pueden emerger de modo automático por superposición de los trazos de memoria correspondientes a experiencias específicas.

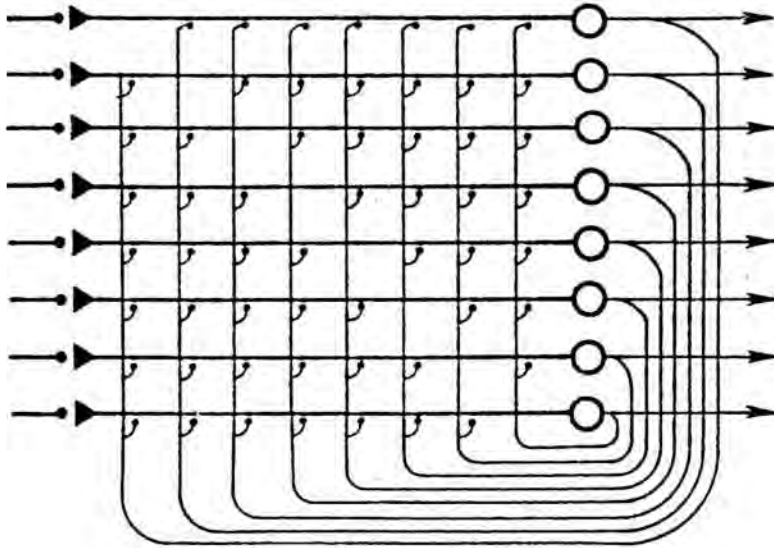
Como se ha señalado repetidamente, los modelos computacionales-discretos poseen un carácter fundamentalmente "abstractivo", en el sentido de que postulan como fundamental la existencia de un lenguaje abstracto de la mente distinto de los "lenguajes" particulares o modales asociados, por ejemplo, a los distintos canales sensoriales. Como señala Rivière (1986), una de las dificultades básicas de este tipo de planteamiento es un fuerte "ceguera" ante los aspectos contextuales de la información; en el polo opuesto, la teoría de código dual sería criticable precisamente por su excesivo contextualismo. Pensamos que los modelos distribucionales permiten dar una respuesta adecuada a esta

problemática, puesto que son capaces, al propio tiempo, de abstraer los caracteres comunes a un conjunto de experiencias y de mantener, en buena parte, su especificidad. En este sentido, este tipo de modelos nos parecen también preferibles, por ejemplo, a los modelos enumerativos (Jacoby, 1983a, 1983b) los cuales, además de exigir una capacidad de almacenamiento virtualmente ilimitada, postulan que los conceptos abstractos no son sino conveniencias notacionales para el psicólogo; para este tipo de modelos, cada suceso sería almacenado en forma de una representación extremadamente rica y estructurada, pero no abstracta. Aunque los modelos enumerativos permiten superar algunas inconsistencias internas de aproximaciones anteriores (por ejemplo, el hecho de que algunos modelos presuntamente no abstractivos postulen que cada ocurrencia de un cierto estímulo -por ejemplo, una palabra- produzca cambios en la activación del mismo detector), McClelland y Rumelhart (1985) demuestran que la aproximación enumerativa no es capaz de dar cuenta de un cierto número de datos empíricos importantes que parecen exigir inequívocamente la presencia de procesos abstractivos "reales", y no meramente como elementos heurísticos.

Si se desea formular una alternativa coherente a los modelos abstractivos clásicos resulta fundamental que la dinámica del sistema distribucional que se postula sea capaz de aprehender la estructura del entorno (siempre en términos de patrones de activación producidos por los inputs sensoriales) de forma que su comportamiento futuro presente un

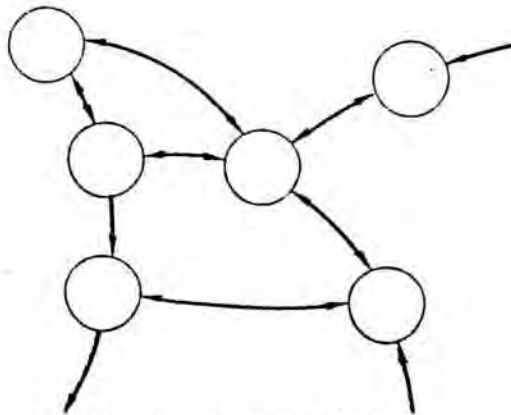
conjunto de regularidades apropiadas para la acción del sujeto en ese mismo entorno. Posiblemente la forma más adecuada de aproximarse a este objetivo sea postular una estructura de unidades elementales organizadas en módulos (McClelland y Rumelhart (1985).

La estructura básica de un módulo se presenta en la figura 4.3. Como se observa en dicha figura, las unidades de un módulo presentan tanto conexiones con otros módulos como inter-conexiones internas modificables. Entonces, si consideramos los estados mentales como patrones de activación de las unidades de un cierto subconjunto de módulos, los trazos de memoria pueden ser introducidos en el modelo en forma de cambios en los pesos de las conexiones establecidas entre las unidades. Aunque el trazo de memoria no es una copia directa del estímulo presentado, éste puede ser reconstruido a partir de aquel. Más aún, este tipo de estructura modular capacita al sistema para reconstruir un ítem de memoria completo a partir de la presentación de una parte del mismo. Deseamos señalar, además, que este tipo de proceso reconstructivo justifica la extrapolación al ámbito de la memoria de la ley gestaltista del cierre, lo cual puede ser un buen indicio de una deseable tendencia hacia la unificación de distintas aproximaciones al sistema cognitivo que poseen en común su carácter dinámico (en oposición al carácter cuasi-estático que hemos atribuido a la aproximación computacional-discreta).



(a) Estructura interna de un módulo

Figura 4.3 Representación del sistema de conexiones intra y entre módulos



(b) Conexiones entre módulos

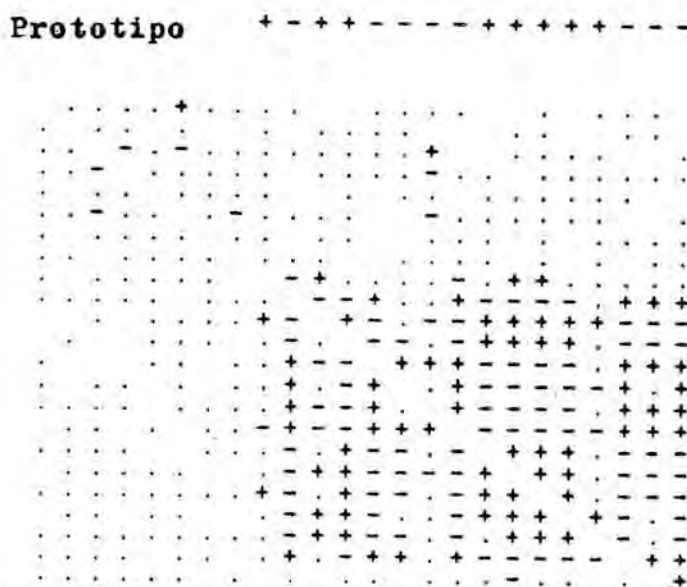


Si atribuimos a la estructura modular algunas de las propiedades de los procesos en cascada que analizábamos anteriormente podremos demostrar de forma inmediata la capacidad abstractiva del sistema. Supongamos, por ejemplo, que se trata de almacenar dos ítems en un módulo único compuesto por 8 unidades. Si consideramos, en aras de la simplificación, que inicialmente los pesos conexionales entre las unidades son nulos, resulta claro que la primera presentación de un ítem dará lugar a un patrón de activación de bajo nivel "resonante", es decir, las distintas unidades responderán débilmente en la dirección inducida por el estímulo. Esta respuesta se irá fortaleciendo a medida que el estímulo sea presentado repetidamente (es decir, en los términos que introduciremos en el siguiente apartado, estímulo y patrón de activación entrarán progresivamente en un régimen de funcionamiento común). Lo más importante, sin embargo, es mostrar cómo un segundo ítem puede ser almacenado por el mismo módulo mediante el establecimiento de un conjunto adecuado de pesos interconexionales. Este proceso se muestra detalladamente en la tabla 2, a la cual nos remitimos para su explicación. Debemos señalar, en todo caso, que el elemento crucial del modelo, que permite el almacenamiento de diversos ítems en un conjunto único de unidades o la reconstrucción de un elemento a partir de la presentación de una parte del mismo, es la denominada regla Delta, una variante continua del procedimiento de convergencia propio de las unidades tipo perceptrón que, además, ha sido propuesta en

Tabla 3

	Input o respuesta para cada unidad							
Patrón 1	+	-	+	-	+	+	-	-
Respuesta inicial	+5	-5	+5	-5	+5	+5	-5	-5
Idem después ensayos	+7	-7	+7	-7	+7	+7	-7	-7
Item incompleto	+	-	+	-	+	+	-	-
	+6	-6	+6	-6	+4	+4	-4	-4
Distorsión	+	-	+	-	+	+	-	+
	+6	-6	+6	-6	+6	+6	-6	+1
Patrón 2	+	+	-	-	-	+	-	+
respuesta inicial	+5	+5	-5	-5	-5	+5	-5	+5
Resp. 10 ensayos	+7	+7	-7	-7	-7	+7	-7	+7
Retest	+7	-7	+7	-7	+7	+7	-7	-7

(a) Comportamiento de un módulo de 8 unidades ante la presentación de 2 items inicialmente, después de 10 ensayos, de forma incompleta o distorsionada



(b) Pesos adquiridos en el aprendizaje de ejemplos distorsionados de un prototipo. Los puntos o espacios en blanco corresponden a pesos de valor muy bajo. El módulo contiene un total de 24 unidades.

algunos modelos de redes adaptativas (Sutton y Barto, 1981). La regla Delta ha sido ideada para determinar el tamaño y dirección del cambio para cada una de las unidades implicadas. La idea básica es la siguiente: para reconstruir un patrón a partir de una parte del mismo, es necesario que el efecto del input interno (procedente de las demás unidades del módulo) sobre cada unidad sea idéntico al efecto del input externo correspondiente al elemento a reconstruir. La regla Delta se basa en el establecimiento de la diferencia entre inputs internos y externos y en la corrección del hipotético desajuste mediante la modificación de los pesos conexionales entre unidades. La regla Delta permite también determinar los pesos necesarios para el almacenamiento común de varios elementos. Sin embargo, la regla introduce una restricción en el modelo, puesto que exige que el input externo que incide sobre cada unidad sea predecible (o reproducible) a partir de una combinación lineal de las activaciones de las demás unidades (asumiendo siempre el carácter de integrador lineal de las unidades elementales). Pensamos, sin embargo, que esta restricción no es fundamental, en el sentido de que se refiere más a las características de los patrones de activación del sistema que a las de los estímulos externos. En otras palabras, esta restricción no limita en modo alguno el tipo de estímulos que pueden ser almacenados por el sistema, aunque sí introduzca restricciones sobre algunos aspectos del modelo (lo cual, en sí mismo, nos parece más positivo que negativo).

¿Qué podemos decir, en definitiva, sobre los procesos abstractivos?. La idea fundamental se deriva de un modo bastante directo de la exposición que acabamos de realizar: los ejemplos concretos de una categoría dada serán considerados como distorsiones de un prototipo de tal categoría que, a su vez, estará representado en el sistema como el conjunto de pesos conexionales resultantes de la presentación de múltiples ejemplos particulares. Es decir, un módulo (o conjunto de módulos) adquirirá un conjunto de pesos que es continuamente modificado por los nuevos ejemplos de la categoría que sean presentados al sistema; estos cambios continuados garantizan el almacenamiento de las características particulares de los distintos ejemplos, aunque parece necesario postular un proceso de decaimiento para evitar una excesiva complicación del sistema y para explicar, además, por qué los ejemplos presentados más recientemente son evocados con mayor facilidad. En este contexto, resulta claro que los pesos de las unidades que responden a los atributos comunes a la mayoría de los ejemplos de la categoría aumentarán de forma casi continua, de manera que la respuesta del sistema será de distinta intensidad en función de la mayor o menor proximidad de un ejemplo dado al "prototipo".

En lo que se refiere a la cuestión general de las modificaciones que tienen lugar en el sistema cognitivo (como resultado, por ejemplo, del aprendizaje) nuestra exposición será relativamente breve, puesto que, de algún modo, el tipo de transformaciones que hemos expuesto al referirnos a los

modelos de abstracción representan ya un buen ejemplo de la aproximación distribucional a este tipo de problemática. Pensamos que, en términos globales, la cuestión puede ser planteada en los siguientes términos: como hemos señalado en el capítulo 2, uno de los problemas esenciales de los modelos computacionales radicaba en su escasa flexibilidad, que obligaba a prácticamente todas las complicaciones posibles del medio estuvieran ya previstas de algún modo por el sistema cognitivo. Obviamente, este planteamiento implica un considerable grado de innatismo. En el caso de los sistemas de producción, por ejemplo, el sistema debe tener previsto el rango de situaciones a las cuales debe enfrentarse el programa, y debe incluir también un conjunto de reglas pre-establecidas de acción ante estas situaciones (aunque, por ejemplo, algunas de estas acciones puedan consistir en la formación de nuevas reglas como producto de la aplicación de reglas anteriores). Señalábamos entonces el claro peligro de regresión infinita que implicaba este planteamiento. La cuestión general que planteamos en este momento es la de cómo pueden generarse o expandirse dinámicamente redes de unidades como las que estamos considerando. Al referirnos al carácter dinámico de esta generación estamos postulando procesos del mismo tipo que los que hemos estado considerando hasta el momento (es decir, de carácter distribucional, interactivo, paralelo y, al menos parcialmente, continuo), puesto que acudir a mecanismos computacionales del tipo propuesto por los modelos de base discreta implicaría una seria



contradicción y reduciría considerablemente la verosimilitud del enfoque alternativo que estamos intentando delimitar.

En este sentido, pensamos que las aproximaciones fundamentales a la construcción dinámica de redes son la teoría de redes adaptativas de Sutton y Barto (1981), el sistema de conexiones dinámicas de Feldman (1982) y el sistema CID (Connection Information Distributor) de McClelland (1985).

#### 4.3 Propiedades globales de las redes de unidades en paralelo

En el apartado 4.1 de este trabajo señalábamos que uno de los aspectos más interesantes de los modelos distribucionales-interactivos era su enorme potencial emergente. En el apartado 4.2 hemos intentado demostrar como la dinámica de este tipo de sistemas es capaz, al menos en principio, de dar cuenta de fenómenos altamente relevantes para la psicología cognitiva, incluyendo algunos hechos difícilmente tratables desde la perspectiva computacional-discreta predominante en nuestro ámbito de trabajo. Sin embargo, las propiedades de los sistemas distribucionales no se agotan en este punto; como veremos a continuación, la dinámica de este tipo de sistemas parece llevar aparejadas una serie de propiedades de tipo más global u holístico que nos parecen del máximo interés en el desarrollo de la teoría psicológica, y de las cuales, a nuestro entender, los propios autores que trabajan en esta línea no han extraído suficiente partido.

Debemos referirnos, en primer lugar, a la propiedad conocida como resonancia. Es bien sabido que la metáfora de la resonancia ha jugado un papel de cierta importancia en el ámbito de la psicología, aunque casi siempre de forma un tanto inconcreta y con un escaso nivel de desarrollo. Por ejemplo Gibson (1966) utiliza el término "resonancia" para referirse a los supuestos mecanismos internos que siguen a la percepción. De hecho, Gibson no se refiere propiamente a "resonancia de energía" sino a "resonancia de información", incidiendo en un cierto paralelismo que concuerda muy bien con lo que exponíamos en el capítulo anterior. Este tipo de argumentos ha sido recogido, por ejemplo, por Shepard (1984) quien, en su análisis de las representaciones internas, muestra que la metáfora de la resonancia permite incluir en los modelos de la representación ciertas propiedades que pueden ser caracterizadas como "restricciones ecológicas" introducidas a lo largo de la filogénesis. Aunque, como acabamos de señalar, Gibson se refiere básicamente a lo que él denomina "resonancia de información", pueden encontrarse también análisis en un nivel más "energético" como, por ejemplo, las referencias importantes, en el ámbito de la fisiología, a fenómenos de "resonancia neural". Pueden citarse, a este respecto, las ideas de patrones de interferencia (Lashley, 1942), circuitos reverberantes (Ashby, 1954; Rashevsky, 1948), o de asambleas celulares y secuencias de fase reverberantes (Hebb, 1949). Más recientemente, el concepto de resonancia neural ha sido utilizado en análisis más

específicos como, por ejemplo, el que propone Ratliff (1983) en cuanto al funcionamiento de los receptores sensoriales. En un ámbito bastante distinto, Duncker (1945) utiliza el término "resonancia" en su análisis de la solución de problemas desde la perspectiva de la psicología de la Gestalt. Es también relativamente frecuente su uso como metáfora de los procesos de la memoria. Como hemos visto, los teóricos que trabajan en el marco de los modelos distribucionales hacen también algunas referencias a la idea de resonancia. Por ejemplo Hinton (1981), en su trabajo sobre redes semánticas con implantación paralela, señala que el patrón de activación asociado a un ítem complejo cualquiera (por ejemplo, a un "triple") tiene su correspondencia en un estado "resonante" de todo el sistema, dada la existencia de múltiples influencias entre niveles. De hecho, pensamos que puede afirmarse con rotundidad que el carácter "resonante" del funcionamiento de este tipo de sistemas es una propiedad inherente a los mismos que se deriva necesariamente de su carácter altamente interactivo.

En términos generales, es posible demostrar, desde un análisis puramente matemático, la idea fundamental de que la resonancia es una propiedad que emerge de forma natural del funcionamiento de cierto tipo de redes. Greene (1962a, 1962b) muestra que si la información es representada en forma de señales graduadas, entonces incluso una red compuesta solamente por elementos lineales poseerá resonancias características. Este tipo de sistema presentará dos propiedades

fundamentales: a) las "resonancias" que representen información significativa tenderán a ser estables frente a perturbaciones aleatorias; b) configuraciones informativas extremadamente complejas pueden ser representadas a través de un pequeño número de intensidades (es decir, de niveles de activación de unidades elementales o conjuntos de unidades). Es casi innecesario señalar que las dos propiedades son altamente deseables, y que ambas pueden ser consideradas como características fundamentales de los modelos distribucionales interactivos.

Más recientemente, Grossberg (1980) ha retomado el análisis de Greene para sistemas no lineales y ha demostrado la relevancia del fenómeno de resonancia en este tipo de redes.

Generalmente, la noción de resonancia no es analizada matemáticamente de forma explícita más que en el caso muy particular de un oscilador lineal sometido a un impulso de frecuencia igual a su frecuencia propia. Sin embargo, incluso en este caso concreto un análisis sencillo de los estados de equilibrio del sistema y de su dinámica general (descrita por una ecuación relativamente simple) revela la existencia de regiones en las cuales la función que describe el sistema puede tomar hasta tres valores (indicando, de este modo, la presencia de lo que en el capítulo anterior denominábamos "bifurcaciones") (Stewart y Peregoy, 1983). Resulta evidente, entonces, que incluso los sistemas resonantes más sencillos muestran un comportamiento en el cual existe la posibilidad de cambios cualitativos importantes que, a nuestro

entender, deben ser analizados y formar parte de cualquier modelo de carácter mínimamente general.

Consideremos un caso algo menos ideal. Supongamos la existencia de dos diapasones  $D$  y  $D'$ , e imaginemos que  $D$  está vibrando. Si aproximamos  $D$  a  $D'$ , entonces  $D'$  comenzará a vibrar en resonancia con  $D$ , de forma que una parte de la energía cinética de  $D$  se transferirá a  $D'$ . El producto de dos osciladores lineales puede ser analizado en términos topológicos (Peixoto, 1962) tomando como ejemplo de sistema dinámico el de dos círculos sobre cada uno de los cuales se da cierto campo de vectores constante. Puede demostrarse que el sistema resultante no es puramente el producto topológico de los sistemas originales y que, en mayor o menor medida, existirá una degeneración hacia un régimen dinámico común para todo el sistema, el régimen de resonancia. En ciertos casos (matemáticamente, cuando la pendiente de las rectas que definen las trayectorias del campo resultante del producto de los dos sistemas originales es un número irracional), el sistema global presentará resonancia difusa, bastante inestable y fluctuante. En otros casos (cuando la pendiente de las rectas es representada por un número racional), cada elemento perderá totalmente su independencia inicial y ambos entrarán en una dinámica única, el sistema resonante. Expresando en otros términos la idea básica del análisis de sistemas resonantes, dos sistemas dinámicos sólo pueden intercambiar energía por resonancia si presentan modos vibratorios que posean características cualitativas



comunes. Estas características admiten un análisis basado en conceptos topológicos. Esta idea se deriva, en parte, de la mecánica hamiltoniana, que reconduce el estudio de los movimientos de ciertos sistemas de puntos al estudio geométrico del correspondiente espacio de fases. En el caso del oscilador lineal, el formalismo hamiltoniano permite transformar el sistema introduciendo una variable oculta  $p$  que nos da el momento. Entonces, la ecuación diferencial inicial que define el movimiento del oscilador ( $q' = -kq$ ) puede ser substituída por una hamiltoniana cuadrática que tomaría la siguiente forma:

$$H = \frac{1}{2} \frac{p^2}{m} + \frac{1}{2} k^2 q^2$$

donde  $m$  es la masa del oscilador. Las trayectorias pueden interpretarse entonces como elipses alrededor del origen. Se da un mínimo para la hamiltoniana  $H$  en el origen  $p = q = 0$  y, generalmente, para  $H = E$  (energía del sistema). Puede afirmarse, entonces, que el sistema oscilante posee una interpretación morfológica en términos de las singularidades de una hamiltoniana. Como veremos en el siguiente capítulo, este tipo de planteamientos son básicos en la formulación de la teoría de las catástrofes.

Entonces, para una energía global  $E$ , puede admitirse que las características vibratorias de un sistema  $S$  están enteramente definidas por una figura  $T(e)$  (función de  $E$ ) en un cierto espacio euclidiano: esta figura puede ser llamada espectro del sistema. Cuando dos sistemas,  $S$  y  $S'$ , entran en interacción, existe una identificación entre los dos espacios

asociados a ellos; en la medida en que los dos espectros puedan ser superpuestos existirá la posibilidad de interacción resonante entre ellos. La resonancia será aguda si, para un conjunto de pares de valores  $E-E'$  tales que  $E + E' = h$  (energía total de los dos sistemas), los dos espectros pueden ser exactamente superpuestos. Obviamente, el caso extremo será aquel en que los dos sistemas sean estrictamente idénticos, de forma que  $E = E' = h/2$ .

Pensamos que este tipo de análisis posee implicaciones importantes para la temática que estamos tratando. Como se señalaba en el capítulo 2 de este trabajo, uno de los problemas básicos (si no el fundamental) de la psicología cognitiva radica en la necesidad de definir las propiedades de un supuesto nivel abstracto de representación que permitiera superar los problemas planteados a las teorías estrictamente modales. Señalábamos también entonces que el formalismo proposicional (y el modo de procesamiento discreto y secuencial asociado a él) presentaba, a nuestro entender, dificultades inaceptables en el momento en que intentábamos descender desde el nivel competencial al terreno algorítmico subyacente a la ejecución. Hemos planteado también la posibilidad de recurrir a un análisis en términos de patrones de activación que se darían sobre una arquitectura de tipo paralelo y con propiedades de procesamiento de tipo continuo. Sin embargo, como señalábamos en el apartado inmediatamente anterior, quedaba por concretar la idea fundamental de semejanza entre patrones de activación. ¿En qué radica esa

semejanza?. En este aspecto la aportación de la metáfora de la resonancia nos parece fundamental. En efecto, si puede demostrarse que los sistemas dinámicos del tipo propuesto en los modelos distribucionales incorporan como característica global la presencia de dinámicas resonantes, y si este tipo de dinámicas es analizable en los términos que acabamos de exponer, puede derivarse de modo inmediato la idea de que los fenómenos de resonancia introducen la posibilidad de una descripción en términos topológicos de las similitudes cualitativas entre patrones de activación y, además, refuerzan una concepción global basada en la suposición de que el sistema cognitivo humano se comporta en gran medida en base a procesos de tipo continuo que incorporan algunas de las propiedades que estamos considerando.

Este tipo de descripción plantearía un conjunto de ventajas que nos parecen fundamentales. Citaremos, de momento, dos de ellas: en primer lugar, en un análisis de este tipo podemos prescindir de algunas restricciones cuantitativas muy perturbadoras. Por ejemplo, no será necesario hablar de semejanza entre patrones en términos de niveles de activación cuantitativamente próximos. En segundo lugar, el punto de vista adoptado por las teorías de la estabilidad estructural nos permite afirmar, al menos en principio, que la dinámica de un sistema basado en conexiones continuas no está en contradicción con la presencia de fenómenos cualitativamente distintos y de discontinuidades altamente significativas. Estas discontinuidades pueden abarcar desde fenómenos tipo

umbral, básicos en psicofísica e incluidos como elementos hipotéticos en muchos modelos distribucionales interactivos, hasta cambios cualitativos de tipo mucho más global como la aparición de discontinuidades en el aprendizaje (Kolers y Duchnicky, 1985), de estadios de desarrollo cualitativamente distintos en sistemas evolutivos (Molenaar, 1986; Molenaar y Openheimer, 1985) o de transformaciones repentinas en la solución de problemas, como proponía clásicamente la psicología de la Gestalt (Köhler, 1947).

Consideremos, por ejemplo, el concepto de re-estructuración propio de la psicología de la Gestalt, y su correlato fenomenológico, es decir, la experiencia de "comprensión súbita" o "insight". Clásicamente, los psicólogos gestaltistas han insistido en dos características fundamentales de la comprensión súbita que nos parecen fundamentales: por un lado, la comprensión súbita de la estructura de un problema (y, en general, su solución inmediata) serían el resultado de un proceso dinámico subyacente en el cual juega un papel fundamental la idea de tensión del sistema. Esta tensión, creada por la presentación de una situación-problema, produciría un conjunto de transformaciones, de "reequilibrios de fuerzas" en el campo psicológico, que en un momento dado se traducirían en la experiencia de insight. Por otra parte, esta experiencia tendrá un carácter repentino y frecuentemente inesperado, puesto que es el resultado de transformaciones reguladas por elementos tensionales que no pertenecen al ámbito de la conciencia.

Pensamos que este tipo de planteamientos, a pesar de su reconocida vaguedad, resultan muy coherentes con el análisis que estamos realizando. Más aún, los sistemas distribucionales interactivos parecen ofrecer un marco excelente para el estudio de este tipo de fenómenos. En el siguiente capítulo intentaremos mostrar, en términos generales, cómo la introducción de una teoría de la estabilidad estructural permite una aproximación interesante a esta problemática.

No obstante, consideramos necesaria una referencia previa a algunas discusiones recientes referidas precisamente a algunos conceptos gestaltistas, especialmente los de fijación e insight. Nos referimos esencialmente a la polémica entablada en torno a la utilidad y relevancia de estos conceptos para una teoría del pensamiento (Weisberg y Alba, 1981a, 1981b, 1982; Dominowski, 1981; Ellen, 1982; Chin-Tung Lung y Dominowski, 1985). Pensamos que este tipo de discusiones clarifica el estatus de este tipo de propiedades o fenómenos globales asociados a sistemas dinámicos y, en este sentido, nos parecen fundamentales para su adecuada modelización.

Weisberg y Alba (1981) plantean un conjunto de experimentos realizados a partir de la conocida "tarea de los 9 puntos", y que, de acuerdo con los propios autores, crean graves dificultades a la psicología de la Gestalt y parecen abogar en favor de una aproximación basada en la "teoría de las hipótesis" (Levine, 1975). Estos experimentos han dado lugar a una amplia controversia en las páginas del "Journal of



Experimental Psychology", y a ellos nos referiremos a continuación.

El primer experimento de Weisberg y Alba señala que el hecho de proporcionar al sujeto una indicación verbal del tipo "los 9 puntos no constituyen un cuadrado" no produce sino una mejoría leve y no inmediata en la ejecución de los sujetos. Los autores sugieren, a partir de este dato, que el "insight" supuestamente producido por dicha indicación no parece suficiente para conducir a una solución rápida y efectiva, tal y como podría esperarse a partir del punto de vista gestaltista. A este respecto, es necesario señalar que los autores consideran, a nuestro juicio incorrectamente, que una simple información verbal ("los 9 puntos no constituyen un cuadrado") debe ser automáticamente capaz de producir un fenómeno complejo como es el de la re-estructuración perceptiva y la consiguiente ruptura de la fijación en una estructura perceptiva errónea (la consideración de 9 puntos como una figura cuadrada). ¿Debe suponerse que una indicación verbal posee la capacidad de inducir de forma inmediata un fenómeno que, probablemente, implica procesos de naturaleza bastante distinta?

Los datos de Weisberg y Alba muestran, por otra parte, que el simple hecho de que el experimentador trace sobre los 9 puntos una sola línea que supere los límites del "cuadrado" permite al 62 % de los sujetos resolver el problema en los siguientes 5 intentos, por término medio (es decir, posiblemente tras un simple proceso de ensayo-error destinado a

encontrar la solución correcta dentro del nuevo ámbito de actuación definido por las soluciones que incluyen líneas que superan los límites del "cuadrado"). Los autores indican, a este respecto, que si en lugar de una sola línea se le da al sujeto orientación sobre las dos primeras líneas de la solución, la proporción de sujetos que resuelve el problema en los siguientes 5 ensayos llega hasta el 100 %, con una media de ensayos suplementarios de 1.44. Evidentemente este resultado es trivial, dado que lo que se ha conseguido es únicamente reducir el ámbito de exploración del sujeto una vez éste ha hallado el dominio correcto de solución. Al parecer, el hecho de que dicho dominio no se componga de una única solución es visto por Weisberg y Alba como una dificultad para la teoría de la Gestalt. Creemos que, en todo caso, se trata de un ejemplo evidente de confluencia entre dos ámbitos teóricos distintos, pero no necesariamente opuestos. Parece perfectamente verosímil considerar que la re-estructuración, en el sentido gestaltista del término, es el hecho que permite la transición de un dominio de solución infructuoso a otro en el cual el problema podría ser resuelto. En este sentido, el hecho de trazar una línea que supere los límites del "cuadrado" puede ser considerado como un indicador mucho más directo para la ruptura de la fijación que la formulación verbal de la misma idea.

En todo caso, es difícil entender cómo el hecho de que el dominio correcto de solución contenga más de un resultado posible pueda esgrimirse como argumento contra la existencia

de un proceso de transición entre uno y otro dominio. Resulta particularmente interesante que, en plena crítica de la aproximación gestaltista a la solución de problemas, Weisberg y Alba señalen que los datos indican que "el problema de los 9 puntos es difícil a causa de que el dominio que contiene la solución correcta esencialmente no existe para muchos sujetos". Difícilmente podrá expresarse de forma más adecuada el efecto básico de la fijación perceptual en este tipo de problemas.

Por otra parte, los datos experimentales obtenidos a lo largo de la polémica parecen indicar claramente que la transición de un dominio de solución a otro no se produce después de agotar todas las soluciones del primer dominio, como proponen Weisberg y Alba, sino que puede tener lugar casi en cualquier momento del proceso. Este dato, junto con los reportes de los propios sujetos experimentales, parecen apoyar una interpretación en términos de ruptura repentina de la fijación.

Debemos señalar, además, que el carácter inesperado de la experiencia de "insight" se pone de manifiesto en el trabajo realizado por Metcalfe (1986). Esta autora ha demostrado que un conjunto de sujetos experimentales que eran capaces de hacer predicciones bastante acertadas sobre su propio rendimiento en tareas de memoria no podían, en cambio, realizar predicciones significativas sobre su rendimiento en una serie de problemas de "insight" de los cuales desconocían la respuesta. Los datos son interpretados por la autora como

indicación de que los problemas de insight implican una "iluminación súbita", y que esta "iluminación" no puede ser predicha con anticipación.

5. Hacia una teoría de la estabilidad  
estructural en psicología

---



## 5.1 La teoría de las catástrofes como análisis de la estabilidad estructural: Elementos básicos

Permitásenos comenzar este apartado citando de forma inmediata un problema importante, a saber, el referido a la "naturalidad" de la teoría de las catástrofes. La propia denominación de este marco conceptual nos parece altamente desafortunada en un doble sentido: En primer lugar, y como aspecto menos relevante, por la existencia de ciertas connotaciones negativas asociadas con el término "catástrofe" que no corresponden en absoluto al concepto fundamental de "bifurcación" o "salto de un dominio de ecuaciones diferenciales a otro". Solamente podrá hablarse de "catástrofe" en este sentido en el caso particular de puntos críticos que marquen un límite a partir del cual se produzca la destrucción del sistema. En segundo lugar, y como idea fundamental, debe citarse el hecho de que los propios creadores de la teoría le atribuyen un carácter más metodológico que teórico (Thom, 1972, 1980b/1985; Isnard y Zeeman, 1976). En un sentido parecido, Huckfeldt, Kohfeld y Likens (1982) se refieren explícitamente al análisis de la "geometría de la estabilidad" como "tecnología matemática", en el sentido de que suministra un conjunto de técnicas de investigación destinadas a aprehender la dinámica de un sistema en términos de sus estados de equilibrio, discontinuidades, etc. La idea básica es la siguiente: la teoría de las catástrofes proporciona un conjunto de técnicas de interpretación que intentan

elaborar el "objeto matemático" más simple que permita generar un conjunto dado de datos empíricos. En otros términos, como señala Thom (1980/1985), la teoría de las catástrofes no es una teoría científica en el sentido habitual del término; en particular, no puede ser contrastada por la experiencia, de la misma forma en que no pueden serlo los instrumentos matemáticos en sí mismos. En este sentido, el cálculo diferencial, por ejemplo, tampoco puede ser "contrastado empíricamente" como tal, aunque sí puedan serlo los modelos de la realidad construidos a partir de este instrumento matemático. En tanto que conjunto de técnicas matemáticas, lo único que puede exigirse a la teoría de las catástrofes es un adecuado nivel de rigor matemático, y en este sentido, a pesar de ciertas críticas debidas, sobre todo, a Sussman y Zahler (1978), parecen existir muy pocas dudas. Volveremos brevemente sobre esta cuestión algo más adelante.

Retomando el concepto de cálculo diferencial, es bien sabido que este instrumento matemático fue ideado para describir la evolución de los estados de un sistema; si, por ejemplo, pretendemos describir los sucesivos estados de movimiento de un cuerpo material, nos bastará con hallar una ecuación diferencial que describa la trayectoria del móvil y que permita, para cualquier momento dado  $t$ , establecer la posición del objeto. Podemos hablar entonces, en un sentido muy real, de un "espacio de fases" o conjunto de estados posibles del sistema, que vendrán dados por las características de la ecuación diferencial que describa el movimiento. La

pretensión de la teoría de las catástrofes es, al menos en parte, la de explorar este espacio de fases a fin de valorar la posible existencia de un conjunto de puntos (los puntos de catástrofe) en cuya vecindad el sistema sufre un cambio cualitativo que se traducirá en la presencia de una discontinuidad en su evolución temporal; esta discontinuidad llevará aparejado el paso de un sistema diferencial a otro.

Naturalmente, no todos los sistemas presentan discontinuidades en su comportamiento (al menos, dentro del rango de valores que puede ser razonablemente estudiado). Un primer elemento a considerar son las condiciones o propiedades que definirán un proceso como discontinuo y, en consecuencia, posibilitarán una aproximación basada en la teoría de las catástrofes. Supongamos, por ejemplo, un caso particular que presente las siguientes características:

- . La conducta es bimodal para algunos valores de las variables que controlan el proceso (en otros términos, a un valor dado de cierta variable antecedente le corresponde más de un valor de una variable consecuente)
- . Se observan cambios súbitos desde una de las modalidades de comportamiento observadas a la otra
- . Se produce el fenómeno de histéresis, es decir, el cambio desde una modalidad de comportamiento a otra se produce para valores distintos de las variables de control en función de la dirección del cambio (es el caso, por ejemplo, de la utilización de ciertos métodos psicofísicos para la determinación de umbrales

sensoriales)

- . Existe una zona de comportamiento inaccesible para ciertos valores de las variables de control
- . Queda implicada la posibilidad de conducta divergente

En la figura 5.1 se presenta un caso particular que cumple algunas de estas características. En principio, cuando el análisis de la evolución temporal de un sistema y de su nube de puntos característica presenten algunas de estas propiedades será posible hablar de la presencia de discontinuidades en el proceso y abordar su estudio desde la perspectiva "catastrofista". En este caso particular, probablemente el modelo más adecuado sea el que corresponde a la "catástrofe de cúspide"; a partir de esta idea, es factible elaborar un modelo que describa adecuadamente la evolución del sistema. Aunque volveremos más adelante sobre el tema, es necesario recalcar aquí que la teoría de las catástrofes no se limita a la aplicación de "recetas diagnósticas" del tipo que acabamos de describir, aunque ésta quizá sea una de las actividades más frecuentes de los investigadores que trabajan en este ámbito; en realidad, este tipo de recetas es parte de la teoría de las catástrofes elementales la cual, a su vez, no es sino un subconjunto específico de la teoría de las catástrofes.

A partir de lo expuesto hasta el momento es fácil llegar, siquiera intuitivamente, a la idea de aplicación como descripción de la correspondencia entre un conjunto de inputs y outputs del sistema; en particular, la teoría de las

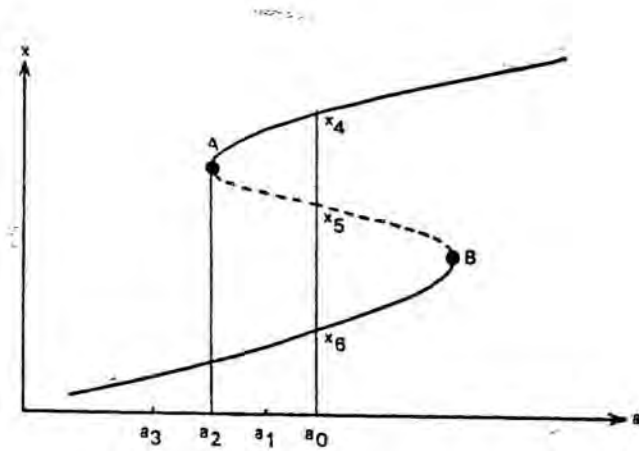


Figura 5.1 Ejemplo de situación en la cual  $a_2$  un valor de la v.I le corresponde más de un valor de la v.D.

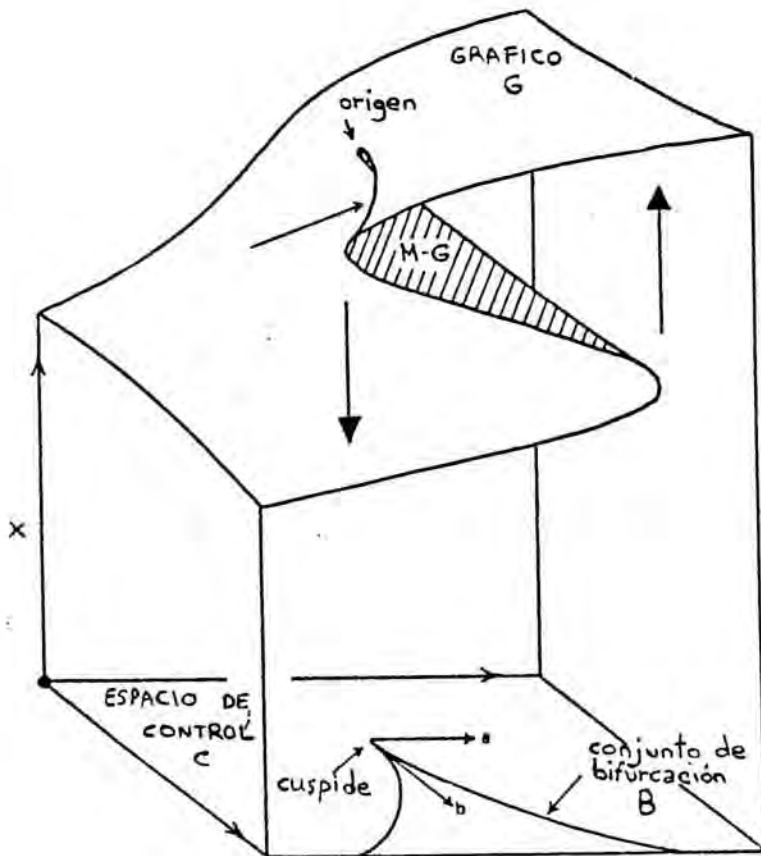


Figura 5.2 Catástrofe de cúspide



catástrofes parte de la consideración de las aplicaciones en las cuales a una cierta entrada le corresponde un número finito de salidas (lo cual implica ya una considerable simplificación). Resulta casi innecesario decir que existe un tipo de aplicación aún más particular en el cual a una entrada cualquiera le corresponde una sola y única salida (siempre, naturalmente, que nos movamos en el dominio de existencia la función); en consecuencia, en este tipo de aplicaciones la evolución del sistema viene determinada por una única función. Las aplicaciones propias de la teoría de las catástrofes pueden ser consideradas, en este contexto, como conjuntos o "paquetes" de funciones.

En términos algo más técnicos, y siguiendo parcialmente el análisis clásico realizado por Thom (1972), la idea que acabamos de exponer podría ser expresada del siguiente modo: Supongamos que el estado de cierto sistema es caracterizado de forma completa por un conjunto de  $q$  parámetros ( $X_1, X_2, \dots, X_q$ ); obviamente, si los valores de cada uno de estos parámetros se sitúan en un eje, el resultado es un espacio euclidiano de orden  $R_q$ . Es posible entonces definir la distancia euclidiana entre dos puntos del espacio  $R_q$  de acuerdo con la expresión:  $d(x,y)^2 = \sum_{j=1}^q (X_j - Y_j)^2$ . El conjunto de puntos situados a una distancia de  $X$  inferior a un número real  $R$  puede ser representado como una esfera de centro  $X$  y de radio  $R$ .

Para cualquier espacio  $R_q$  puede encontrarse un subconjunto de puntos que representan los estados físicamente

realizables del sistema; llamaremos a este grupo de puntos conjunto de definición del sistema. Dado que los parámetros que regulan el comportamiento del sistema no pueden ser cuantificados con precisión absoluta (por su carácter continuo), siempre se encontrarán, en las cercanías de cualquier punto a que pertenezca al conjunto de definición del sistema, otros puntos que también pertenecerán a ese conjunto (es decir, existirá una esfera de centro a y de radio r que contendrá puntos pertenecientes al conjunto de definición del sistema). Por este motivo, estas esferas serán consideradas como abiertos y, dado que la reunión de una serie de abiertos es un abierto, el conjunto de definición  $M_a$  puede ser considerado también como abierto. El complementario K de este abierto constituirá el cerrado de los puntos de bifurcación, es decir, el conjunto de puntos catastróficos que determinarán un cambio abrupto de la dinámica del sistema. Dado el abierto  $M_a$ , la intersección de todos los cerrados que estén contenidos en  $M_a$  constituye un cerrado  $M'$  denominado adherencia de  $M_a$ . El conjunto  $M' - M_a$  se denomina frontera de  $M_a$ : cerca de cualquier punto c que pertenezca a esta frontera hay puntos que pertenecen a  $M_a$ , pero el propio c no pertenece al conjunto de definición del sistema. Siempre que el punto representativo del sistema llegue a un punto de la frontera se producirá el fenómeno nuevo y repentino que denominamos "catástrofe".

De acuerdo con los términos que acabamos de exponer, el problema clásico de la estabilidad estructural puede quedar

redefinido de la siguiente forma: un sistema será estructuralmente estable (y, por tanto, presentará una morfología reconocible) si el conjunto  $K$  de puntos catastróficos es raro, o "denso en ninguna parte" ("nowhere dense"). Se denomina raro a cualquier conjunto  $B$  de un espacio (métrico o topológico) tal que el interior de su clausura es vacío (siendo la clausura el cerrado resultante de la intersección de todos los cerrados que contiene  $B$ ). Entenderemos por interior de un conjunto en un espacio métrico o topológico la unión de todos los abiertos contenidos en ese conjunto). Expresando la misma idea de forma más asequible (aunque mucho menos precisa), puede decirse, siguiendo a Simmons (1967), que un conjunto raro "no cubre una gran porción del espacio".

Es interesante resaltar, en este punto, que la definición de abiertos y cerrados en un espacio métrico es equivalente a la formulación en términos de entornos propuesta por Hausdorff en 1914 y que permitió la axiomatización de la topología de forma independiente a la noción de métrica. De hecho, como señala Thom (1980/1985), los espacios euclidianos pueden ser considerados como casos particulares de espacios topológicos conexos (un espacio topológico  $X$  se llama conexo si no existen abiertos no vacíos de  $X$  tales que su unión sea  $X$  y su intersección sea el conjunto vacío).

Retornando al planteamiento general, supongamos que los  $q$  elementos del sistema intervienen en algún tipo de proceso. Por ejemplo, los distintos elementos del sistema pueden ser

sustancias químicas (cada una de ellas con su concentración), o las distintas unidades de un sistema distribucional interactivo (cada una de ellas con su correspondiente nivel de activación). En el primer caso, las distintas sustancias pueden intervenir en una serie de reacciones químicas que pueden ser descritas, por ejemplo, mediante los cambios en su balance cinético; en el caso de los modelos distribucionales en psicología, las diversas unidades pueden intervenir en procesos como la representación de un objeto externo, la recuperación de datos de la memoria, o la solución de un determinado problema. En cada uno de estos casos, el proceso tendrá como consecuencia cambios sucesivos en los niveles de activación de las distintas unidades, del mismo modo que en el ejemplo anterior se producirían cambios en las concentraciones de las distintas sustancias presentes en las reacciones. En ambos casos podemos, en principio, obtener una expresión que relacione los cambios en los parámetros asociados a los distintos elementos en función de los valores anteriores de esos mismos parámetros. En el ejemplo de las reacciones químicas, puede obtenerse una expresión como la que sigue:

$$(E) \quad d C_i / d T = X_i (C_j)$$

siendo  $C_i$  la concentración de la sustancia  $i$  y  $T$  el tiempo. Obviamente, en el caso de la activación de unidades en un sistema distribucional podrían derivarse expresiones semejantes. Las ecuaciones  $E$  definirán sobre el dominio  $D$  (o conjunto de definición) de los estados posibles del sistema

un campo de vectores de componentes  $X_1, X_2, \dots, X_q$ . Si las funciones  $X_1, \dots, X_q$  son regulares se puede integrar (al menos localmente) este sistema y, en consecuencia, derivar leyes del tipo  $C_i(t) = H_i(t, C_{i0})$ , siendo  $C_i(t)$  la concentración de la substancia  $i$  en el momento  $t$  y  $C_{i0}$  la concentración en el momento  $t = 0$ . Fijadas las constantes  $C_{i0}$ , las funciones  $H_i$  definen sobre el dominio  $D$  una curva diferenciable que describe la evolución del sistema a lo largo del tiempo.

De la exposición anterior se derivan algunas ideas generales sobre el tipo de modelos que pueden construirse desde el enfoque que estamos considerando. En particular, es fundamental resaltar que estos modelos deberán incluir dos aspectos distintos: el "espacio de fase" o campo de estados posibles del sistema, y la evolución temporal del mismo. El estudio de los espacios de fase es el objeto de la **geometría diferencial**; el estudio de las trayectorias definidas en esos espacios es objeto de la **Dinámica**. Las ideas fundamentales en geometría diferencial serán las de **aplicación diferenciable** y **variedad diferenciable**; en el campo de la dinámica, los conceptos básicos son los de sistema dinámico, estados asintóticos y, muy especialmente, la idea de **atractor**.

Muy brevemente, diremos que una aplicación definida por las fórmulas  $Y_k = F_k(X_1, X_2, \dots, X_n)$  es continuamente diferenciable  $r$  veces (o bien que es de clase  $C_r$ ) si las funciones  $F_k$  tienen derivadas parciales continuas hasta el orden  $r$  (incluido). En el marco que estamos considerando,



las aplicaciones se establecerán entre un espacio-origen (que corresponderá al espacio de los parámetros de control del sistema) y un espacio-destino que, generalmente, será el espacio de las conductas.

Sea  $F: \mathbb{R}^n \rightarrow \mathbb{R}^{n+k}$  una aplicación inyectiva que en todo punto del conjunto origen es de rango  $n$ ; el conjunto imagen  $F(\mathbb{R}^n)$  será denominado entonces variedad "sumergida" de dimensión  $n$  en  $\mathbb{R}^{n+k}$  (o también variedad de codimensión  $k$ ).

Igualmente, definiremos variedad diferenciable  $V$  de orden  $k$  como un espacio que puede ser representado localmente por un sub-conjunto cerrado del espacio euclidiano  $\mathbb{R}^{n+k}$ , y tal que en todo punto  $x$  de este conjunto existe una representación local de  $V$  a través de un sistema de  $k$  ecuaciones  $F_1 = F_2 = F_3 = \dots = F_k = 0$ , donde las diferenciales  $dF_1 \dots dF_k$  son linealmente independientes en  $x$ .

De esta definición resulta que en cada punto  $x$  de la variedad podemos encontrar un "mapa local" en la que  $V$  es una variedad lineal de dimensión  $n$  (o de codimensión  $k$ ). En otros términos, la variedad diferenciable puede ser definida como un espacio topológico recubierto por "mapas" locales  $U$ ; cada  $U$  es un abierto, y en la intersección (no vacía) de dos abiertos el mismo punto del espacio tiene dos representantes, cada uno en uno de los dos abiertos; se postula entonces que la correspondencia entre los dos puntos viene dada por un difeomorfismo (en otras palabras, un cambio de coordenadas curvilíneo).

Consideremos el conjunto de las funciones de  $R^q$  sobre  $R^m$  continuas en un entorno de un punto  $x$ ; se podrá definir entonces una relación de equivalencia de la forma siguiente:  $f$  y  $g$  serán equivalentes si y sólo si existe un entorno de  $x$  en que  $f$  y  $g$  coinciden. La clase de equivalencia para esta relación (es decir, el conjunto de las funciones  $g$  tales que  $g \sim f$ ) se denominará germen de la función  $f$ . Sean entonces dos aplicaciones locales del espacio  $R^q$  sobre  $R^m$ ,  $f$  y  $g$ , que aplican el origen  $O$  del espacio-fuente sobre el origen  $O'$  del espacio-destino; si estas dos aplicaciones tienen las mismas derivadas parciales hasta el orden  $r$  (incluido), podrá afirmarse que ambas poseen el mismo germen local de orden  $r$  en  $O$ . El conjunto de los gérmenes forma un espacio vectorial  $J_r(q,m)$  en el cual existe un subconjunto  $K_r$  de bifurcación que posee la siguiente propiedad: si el germen de orden  $r$  de dos aplicaciones  $f$  y  $g$  es el mismo y es un punto del conjunto complementario  $J_r(q,m) - K_r$ , estas dos aplicaciones son localmente del mismo tipo topológico; existirán entonces homeomorfismos  $h, h'$  del conjunto origen al conjunto destino tales que el diagrama siguiente es conmutativo:

$$\begin{array}{ccc} R^n & \xrightarrow{f} & R^p \\ \downarrow h & & \downarrow h' \\ R^n & \xrightarrow{g} & R^p \end{array}$$

con  $h(O) = O$ ,  $h'(O) = O'$ . Entonces, si toda deformación de  $f$  lo bastante pequeña es topológicamente equivalente a  $f$  podrá

hablarse de germen estructuralmente estable. En caso de que no se cumpla tal condición, pueden presentarse dos casos cualitativamente distintos (Thom, 1972): o bien las perturbaciones arbitrariamente pequeñas de  $f$  pueden presentar un número infinito de tipos topológicos (hablaremos entonces de singularidad de codimensión infinita), o bien el número de tipos topológicos es finito (localmente). En este último caso, nos hallamos ante singularidades de codimensión finita. A partir de este punto, la "teoría del despliegue universal" intenta establecer el conjunto de deformaciones topológicamente equivalentes a una función  $f$  que defina una singularidad; en otras palabras, se trata de formular el despliegue universal de la singularidad. Esto resulta posible porque, en el caso de singularidades de codimensión finita  $m$ , es factible "sumergir" la aplicación  $f$  en una familia de deformaciones de  $m$  parámetros tal que la aplicación  $F: R_q \times R_m \rightarrow R_q \times R_m$  es estructuralmente estable. Como caso particular, un germen estructuralmente estable define una singularidad de codimensión cero. Debemos señalar que un despliegue de  $f$  se denomina versal si cualquier despliegue de  $f$  puede ser inducido a partir del germen mediante un morfismo adecuado; si se toma la codimensión más pequeña para la cual un despliegue es versal, entonces ese despliegue puede ser llamado uni-versal.

La aportación fundamental de la teoría de las catástrofes elementales consiste precisamente en analizar el despliegue universal del conjunto de singularidades de codimensión

igual o menor a 4 (es decir, de las singularidades posibles en nuestro espacio-tiempo). El resultado de este análisis es la conocida lista de 7 catástrofes elementales, a las cuales puede atribuirse un carácter canónico, en el sentido de que cualquier singularidad de un germen de codimensión menor de 5 debe ser equivalente a algunas de esas 7 catástrofes. Este es el resultado fundamental del denominado "teorema de Thom" (aunque su propio autor se opone a tal denominación), desarrollado posteriormente por autores como J.N. Mather. En la tabla 3 se presenta el listado de las 7 catástrofes elementales; debe remarcarse, en este punto, que las ecuaciones definitorias de estas catástrofes relacionan el valor de una función de potencial  $P$  (a la cual nos referiremos más adelante) con los valores de la variable de estado del sistema ( $x$ ) y con las variables de control del proceso ( $u, v, \dots$ ). Baste por el momento con señalar que la función potencial es fundamental en el análisis de los sistemas gradientes (a los cuales resulta especialmente aplicable la teoría de las catástrofes), y que la dinámica del sistema vendrá regulada en gran parte por la presencia de singularidades en esa función de potencial. En la figura 5.2 se muestra gráficamente la forma que toma una de las catástrofes elementales.

Desde la perspectiva que estamos considerando, un sistema dinámico  $(M, X)$  es el resultado de una acción diferenciable del grupo  $\mathbb{R}$  de los reales sobre una variedad  $M$ , el espacio de fase. Esa acción define en cada punto de  $M$  un campo de vectores  $X$ . Recíprocamente, a partir de un campo de vectores

	Co-dimensión	Nombre	Centro organizador	Despliegamiento universal
Dimensión uno	0	Simple Mínimo	$V = x^2$	$V = x^2$
	1	El pliegue	$V = x^3/3$	$V = x^3/3 + ux$
	2	La cúspide (Catástrofe de Riemann-Hugoniot)	$V = x^4/4$	$V = x^4/4 + ux^2/2 + vx$
	3	La cola de milano	$V = x^5/5$	$V = x^5/5 + ux^3/3 + vx^2/2 + wx$
Dimensión dos	4	La mariposa	$V = x^6/6$	$V = x^6/6 + tx^4/4 + ux^3/3 + vx^2/2 + wx$
	3	El ombligo hiperbólico	$V = x^3 + y^3$	$V = x^3 + y^3 + wxy - ux - vy$
	3	El ombligo elíptico	$V = x^3 - 3xy^2$	$V = x^3 - 3xy^2 + y(x^2 + y^2) - ux - vy$
	4	El ombligo parabólico	$V = x^2y + y^3$	$V = x^2y + y^3 + wx^2 + ty^2 - ux - vy$

Tabla 3. Listado de las 7 catástrofes elementales



$X$  en  $M$  es posible, por integración local, definir un "germen de acción" de  $R$  en  $M$ . Denominaremos órbita  $g(m)$  de un punto  $m$  de  $M$  al conjunto de las imágenes de  $m$  para las transformaciones  $g$  del grupo  $R$ .

A partir de estos conceptos básicos pueden derivarse las nociones más habitualmente utilizadas en la modelización de la dinámica temporal de un sistema. Sea  $g$  una órbita cualquiera. Un punto  $m$  de  $M$  será denominado punto  $w$ -límite de  $g$  si, para todo  $\epsilon > 0$  y  $t > 0$  existe un tiempo  $T > t$  tal que el punto  $g(t)$  se encuentra a una distancia de  $m$  inferior a  $\epsilon$ . El conjunto de los puntos  $w$ -límite de  $g$ , al que denominaremos  $w(g)$  es un conjunto cerrado invariante de trayectorias. A partir de esta idea podemos definir la noción básica de trayectoria asintótica: dos trayectorias  $g, g'$  son asintóticas si ambas poseen el mismo conjunto  $w$ -límite. Dos puntos  $m, m'$  son asintóticamente equivalentes si sus trayectorias son asintóticas. Esto define entre los puntos de  $M$  una relación de equivalencia. Esto nos permite introducir una segunda definición del objetivo de la teoría que estamos considerando: se trata de precisar la naturaleza topológica de esta relación de equivalencia. En ciertos casos, esta relación puede reflejar una situación determinista; en tales condiciones, a cada estado asintótico le corresponderá un abierto de  $M$ , su "cuenca". En consecuencia, el sistema podrá ser definido completamente mediante un conjunto de trayectorias dirigidas hacia sus correspondientes estados asintóticos. Se trata, obviamente, del caso en que las condiciones iniciales

del sistema determinan de forma absoluta su evolución posterior. Existen, sin embargo, casos en que las distintas clases de equivalencia están inextricablemente "mezcladas", de forma que nos hallamos en una situación de indeterminación. Estamos finalmente en condiciones de definir el concepto, absolutamente crucial, de **atractor**. Se llamará atractor del campo  $X$  definido sobre  $M$  a un subconjunto  $A$  de  $M$  invariante que verifique las condiciones siguientes:

1) Prácticamente toda trayectoria en  $A$  es densa dentro de  $A$   
2) Existe un sistema fundamental de entornos  $U$  de  $A$  dentro de  $M$  tal que:

a.- Toda trayectoria de un punto  $u$  que pertenezca a  $U$  admite  $A$  como conjunto  $w$ -límite

b.- Si un punto  $a$  de  $U$  es tal que el conjunto  $a$ -límite (límite del conjunto límite para  $t \rightarrow -\infty$ ) de  $a$  reencuentra a  $A$ , entonces  $a$  es un punto de  $A$

Tras esta definición técnica se oculta una visión mucho más intuitiva e inmediata. Consideremos un ejemplo general de nuestro campo de estudio: Sea  $X$  un conjunto de estímulos externos; este conjunto es aplicado en el conjunto de los estados del sistema cognitivo  $Y$  mediante una función

$f: X \rightarrow Y$  que puede depender, por ejemplo, de la conexión entre los órganos sensoriales y niveles más centrales (no entraremos ahora a considerar la naturaleza y propiedades de tales conexiones). Podremos suponer entonces que el espacio de los estados cognitivos se encuentra dividido en un determinado número de dominios, cada uno de los cuales está

caracterizado por un campo de vectores que, en un instante dado, es dominado por un cierto atractor. En consecuencia, desde este punto de vista el conjunto de las trayectorias posibles en el espacio queda dividida en una serie de "cuencas", cada una de ellas asociada a un determinado atractor. Es importante señalar que estos atractores pueden estar sometidos a transformación en virtud de las modificaciones inducidas por el entorno del sistema.

Podemos introducir entonces una tercera definición del objetivo de la teoría de las catástrofes: se trata de identificar las condiciones bajo las cuales un cierto atractor puede llegar a ser inestable (convirtiéndose, entonces, en un punto de bifurcación), o bien los cambios que se producen en la evolución del sistema ante la presencia de dos atractores igualmente "potentes" (puntos de conflicto).

El análisis de los sistemas gradientes representa un aspecto fundamental de la teoría de las catástrofes. No resulta sorprendente el énfasis en el estudio de este tipo de sistemas, puesto que la noción de gradiente de potencial introduce una justificación no puramente formal para la idea de atractor. En efecto, en los sistemas gradientes se incorpora un principio de energía mínima (que, como señalaremos oportunamente, juega también un papel importante en psicología) según el cual el sistema tiende a fomentar la transición hacia los estados de nivel energético mínimo de entre los que sean compatibles con una situación dada. Esta transición hacia los estados de energía mínima puede ser descrita muy

adecuadamente en términos de los conceptos básicos de la teoría de las catástrofes. La hipótesis básica es que el campo de vectores  $X(x)$  -que depende diferenciablemente de las coordenadas del punto  $x$  en el espacio de los parámetros del sistema- es, para todo punto  $x$  de  $M$ , un campo de gradiente sobre  $M$ :

$$X(m;x) = - \text{grad } V(m;x)$$

donde  $V(m;x)$  es una función potencial sobre  $M$  que depende diferenciablemente de  $x$ .

En un punto cualquiera  $x$  de  $W$ , los regímenes locales estructuralmente estables son definidos por los atractores estructuralmente estables del campo de vectores  $X$ , es decir, por los puntos donde el potencial  $V(m;x)$  resulta ser mínimo. Dado que existirán diversos mínimos locales de la función potencial, el problema que se plantea a continuación es el de escoger aquel mínimo hacia el cual derivará el sistema. Aunque volveremos sobre esta cuestión en el punto 5.3, debemos señalar aquí que existen dos reglas o convenciones básicas para la toma de esta decisión: la convención de Maxwell, según la cual el sistema derivará siempre hacia el punto de potencial mínimo absoluto, y la "convención de retraso perfecto" ("delay convention"), que dirigirá al sistema hacia el mínimo local más próximo en cada momento dado. Técnicamente, la convención de Maxwell indicaría que, determinado un despliegue de orden  $r$ , el conjunto  $K$  de los puntos de catástrofe consta de los puntos  $u$  del espacio de control  $R_r$  donde la función de potencial alcanza el mínimo absoluto por

lo menos para dos puntos o donde el mínimo es único, pero inestable. Esto corresponde a los dos tipos de atractor que hemos mencionado brevemente unas líneas más arriba.

## 5.2 ¿Matemáticas para la psicología?

Es frecuente la discusión sobre las posibilidades de aplicación de las matemáticas a la psicología. De la misma forma en que se ha considerado que la aplicación del método científico a la psicología no podía dar cuenta de forma adecuada de los procesos implicados en la conducta humana, -y de ahí la clásica distinción entre "Naturwissenschaften" y "Geisteswissenschaften" (Windelband, 1894), o el "reduccionismo fenomenológico" de Dilthey en respuesta a la problemática kantiana- se han formulado también ciertas críticas a la introducción de las matemáticas como instrumento de modelización psicológica.

Se ha señalado, en primer lugar, que los modelos matemáticos podían tener una excesiva potencia y comportar dificultades de cálculo desproporcionadas a su utilidad. A este respecto es necesario responder que, al menos en principio, la dificultad de un instrumento matemático es en buena parte reflejo del objeto de estudio que exige su aplicación. En el caso de la física, por ejemplo, una partícula elemental como el electrón es representada mediante ecuaciones integro-diferenciales de enorme complejidad. Debe señalarse que los modelos matemáticos aplicados a ciertos procesos



psicológicos no han alcanzado todavía el mismo grado de complicación que el modelo básico del electrón. Cabe suponer que el objeto de estudio de la psicología -la conducta humana y sus procesos subyacentes- es más complejo que una "simple" partícula elemental y que, por tanto, presumiblemente la utilización de las matemáticas deberá realizarse a un nivel de complejidad mayor que el actual. Otra cuestión distinta es, como veremos, qué matemáticas serán útiles para la modelización psicológica.

Puede argumentarse también que las matemáticas constituyen un instrumento de modelización excesivamente rígido, y que sus propiedades pueden ser impuestas al objeto de estudio, más que servir para describirlo adecuadamente. Aunque este problema, como hemos afirmado repetidamente en el capítulo 2, es común a muchos instrumentos de modelización (y particularmente al ordenador), debe señalarse que las matemáticas constituyen un sistema de amplias posibilidades y que presenta un grado de flexibilidad considerable. Más aún, e invirtiendo quizá una tendencia histórica, el desarrollo actual de las matemáticas se realiza cada vez más en función de las necesidades de las ciencias fácticas. Se han utilizado instrumentos matemáticos que no habían obtenido previamente un fundamento teórico suficiente. La utilidad y precisión de los cálculos realizados con estas técnicas es, con frecuencia, el origen de la búsqueda de su fundamentación matemática rigurosa.

Un ejemplo de este último punto puede hallarse en el campo de la física. Durante más de 200 años los físicos han utilizado ciertos sistemas de cálculos con infinitésimos que no habían obtenido una comprobación rigurosa en el campo de la teoría matemática. Solamente algunos trabajos recientes (Robinson, 1974; Lutz y Goze, 1981; Harthong, 1984) han desarrollado el análisis no estándar como instrumento de trabajo con infinitésimos, mostrando que la utilidad práctica que el cálculo infinitesimal ha tenido para la ciencia física posee un fundamento matemático preciso.

En su crítica específica a la teoría de las catástrofes, Sussman y Zahler (1978) han utilizado un argumento muy semejante al afirmar que "...la teoría de las catástrofes es un intento de hacer ciencia tratando de imponer al mundo un sistema preconcebido de estructuras matemáticas, más que mediante el método experimental" (pág. 208). En realidad, una lectura detenida del trabajo de Sussman y Zahler muestra que esta crítica parece ir dirigida más a ciertas propuestas realizadas por Zeeman (1977) que al corpus global de la teoría de las catástrofes. La crítica se refiere tanto a cierta falta de rigor en la modelización como al intento de "deducir" fenómenos naturales a partir de teoremas puramente matemáticos. Obviamente, como señala el propio Thom (1980/1985), esto solamente es lícito cuando se ha demostrado que un fenómeno viene regido por un determinado modelo que, a su vez, admite una descripción de tipo matemático.

El mero hecho de aceptar la posibilidad y el interés de la modelización matemática en psicología plantea de inmediato la cuestión del tipo de matemáticas más adecuado para la modelización de los procesos psicológicos. Naturalmente, esto depende tanto de los fenómenos a estudiar como del tipo de modelos que se construyan para ello. Nos parece especialmente pertinente, en este punto, hacer una referencia general al papel de las matemáticas en las ciencias naturales y en las ciencias llamadas "humanas" o sociales.

Es casi innecesario señalar que la física es la disciplina científica matematizada por antonomasia. Sin embargo, el panorama no es tan simple como podría parecer en primera instancia. Es cierto que en física fundamental las grandes leyes clásicas (por ejemplo, la ley de gravitación) permiten construir modelos de extraordinaria exactitud numérica. Sin embargo, tanto en mecánica cuántica (especialmente, en cuanto al problema de las interacciones) como en los niveles más macroscópicos (por ejemplo, física del estado sólido) se encuentran leyes empíricas que no poseen expresión matemática explícita. Como señalábamos en el capítulo 3, en el ámbito de la termodinámica las descripciones son a menudo de tipo bastante cualitativo. Tanto en química como en biología, a pesar del creciente número de trabajos basados en modelos matemáticos, las aplicaciones de las matemáticas suelen reducirse a la elaboración de modelos para situaciones locales (por ejemplo, circulación de la sangre), si exceptuamos algunos aspectos de cinética química, en el primer caso, o

la teoría de poblaciones y la genética, en el segundo. Algo semejante podría decirse de disciplinas como la fisiología, la etología y la propia psicología. En cualquier caso, y al margen del análisis pormenorizado de los desarrollos matemáticos realizados en cada ámbito, parece claro que la utilización de las matemáticas fuera del campo de la física fundamental presenta ciertas dificultades. Pensamos que buena parte de ellas radica en su carácter mimético respecto a los modelos matemáticos cuantitativos propios de la física (o, al menos, de parte de ella).

Thom (1982) realiza un análisis de las razones que explican la gran exactitud de algunas leyes físicas. En términos simples, la cuestión es que los espacios internos que se han introducido para describir las entidades físicas pueden relacionarse directamente con el espacio-tiempo o con su grupo de equivalencias mediante construcciones matemáticamente definidas. La validez y exactitud de las leyes de la física fundamental se basa en el hecho de que, al menos localmente, el espacio-tiempo posee una estructura analítica natural. Esto es cierto desde el momento en que se acepta el punto de vista clásico en física, según el cual el espacio-tiempo es considerado como el espacio homogéneo de un grupo de Lie continuo de equivalencias.

La situación cambia cuando abandonamos el ámbito de las leyes físicas fundamentales. El objetivo sigue siendo el mismo: se trata de encontrar algún medio por el que un dato empírico pueda dotarse de una estructura analítica.

Consideremos el problema de la diferenciación, que conduce a la existencia de especímenes clasificables e identificables. Aquí ya no hay grupo continuo de invariancia, puesto que dos especímenes de un mismo tipo no son necesariamente iguales métricamente; en consecuencia, resulta necesario, frecuentemente, recurrir al espacio de las frecuencias estadísticas de aparición de cada uno de los especímenes. En el ámbito de la química, por ejemplo, las leyes de equilibrio se basan en la regularidad morfológica de los procesos combinatorios entre moléculas que constituyen la base de las reacciones químicas. Los modelos elaborados para este tipo de fenómenos sólo satisfacen la condición de analiticidad si se demuestra que las constantes que figuran en la ley de acción de masas dependen analíticamente de las concentraciones. Sin embargo, nada permite afirmar a priori este tipo de dependencia. En un sentido semejante, la teoría de poblaciones ha podido aplicarse en biología solamente en casos relativamente simples, dado que las constantes que afectan a la frecuencia de las interacciones (por ejemplo, la predación) no permiten una evaluación de tipo analítico. En la mayoría de los casos, la matemática "fuerte" basada en la analiticidad de las funciones implicadas en los modelos debe ser substituída por una matemática "blanda" que permite únicamente, al menos en principio, la derivación de conclusiones de tipo cualitativo relacionadas con los estados asintóticos del sistema (existencia de puntos de equilibrio, ciclos límite, etc.). Por supuesto, el calificativo "blando" se refiere a la



dificultad para la obtención de conclusiones cuantitativamente precisas, y no a la falta de rigor matemático de este tipo de aproximación. Como veremos más adelante, los obstáculos para la obtención de conclusiones cuantitativas constituyen un elemento importante de crítica de los modelos basados en la teoría de las catástrofes.

Existen en psicología antecedentes de la aplicación de este tipo de matemática "débil" o "cualitativa". En realidad, para algunos autores (por ejemplo, los primordios teóricos de la Gestalt) el carácter cualitativo de la investigación psicológica es una característica inherente al estado de desarrollo de la disciplina. Como señala Köhler (1947, 1955), la cuantificación precisa debe ir precedida por un conocimiento preciso de los fenómenos a estudiar, lo cual exige en primer lugar una aproximación de tipo cualitativo. Esto es especialmente difícil en psicología, dada la complejidad de los sistemas y de las relaciones entre sus elementos. Algunas reflexiones semejantes se encuentran en la base de la aproximación topológica propuesta clásicamente por Lewin (1935, 1951) y Brown (1936). Para estos autores, la geometría es de importancia crucial en la modelización psicológica, mientras que la aritmética y el álgebra serían fundamentales para la teoría de la medida. De este modo, el campo psicológico es definido como un espacio en el cual pueden ser expresadas relaciones posicionales. La posición podría ser definida en términos de distancia y dirección en un espacio euclidiano, o en términos de relación en un espacio topológico. Lewin

(1934) señala explícitamente la relevancia de la idea riemanniana según la cual pueden construirse espacios a partir de cualquier dimensión o propiedad; como señalábamos anteriormente, de la formulación de Riemann (1923) se deriva inmediatamente la idea de que las propiedades del espacio substrato de cualquier fenómeno dependen de la dinámica de los procesos que dan lugar a dicho fenómeno. La consecuencia que Lewin extrae de esta idea es que, en términos muy generales, la relación entre las propiedades del espacio psicológico que él postula y la dinámica de los procesos subyacentes es comparable a la que se establece entre propiedades del espacio físico y los correspondientes procesos. Sin embargo, existen importantes diferencias entre los dos ámbitos de trabajo. Por ejemplo, la "locomoción psicológica", entendida como aproximación a una meta, no puede ser aprehendida por la geometría euclidiana utilizada frecuentemente por los físicos. Los procesos psicológicos no pueden "coordinarse unívocamente" con el espacio físico; por ejemplo, una persona puede acercarse a su meta aunque físicamente se aleje de ella. Esta idea es un claro antecedente del análisis, técnicamente mucho más complejo, realizado por Thom (1982).

En el marco de la "psicología topológica" lewiniana la conducta es definida como función del **espacio vital**. Este espacio vital está compuesto por la totalidad de los hechos que determinan la conducta de un individuo en un cierto momento y representa la totalidad de los eventos posibles. El espacio vital incluye a la propia persona y al entorno y puede

ser representado por un espacio finitamente estructurado. La idea de espacio finitamente estructurado corresponde a aquel espacio en el que pueden distinguirse regiones relativamente amplias. En un espacio infinitamente estructurado podrían distinguirse regiones arbitrariamente pequeñas.

Uno de los problemas básicos de Lewin radicaba en el hecho de que, en su tiempo, la topología estaba todavía muy poco desarrollada. De ahí su actitud un tanto especulativa en cuanto a la presentación de nuevas nociones topológicas. Un caso paradigmático de lo que estamos afirmando puede encontrarse en la idea de "espacio hodológico". El propósito del espacio hodológico es el de hallar un tipo de geometría que permita la utilización del concepto de dirección de manera significativa para la psicología (Lewin, 1938). En efecto, Lewin parte de la base de que la "dirección psicológica" no puede ser representada adecuadamente ni por la línea recta euclidiana ni por la "línea más recta" de Riemann. Dado que el espacio psicológico posee el carácter de espacio topológico, Lewin sugiere la necesidad de construir un nuevo tipo de espacio topológico en el cual las direcciones jueguen un papel. Sin embargo, no entra en la formulación matemática de este "espacio hodológico", ni se plantea la posibilidad efectiva de su construcción. Como señala Wolman (1968), "Lewin parece creer en la omnipotencia de las matemáticas y espera sinceramente que algún día se elabore su sistema propuesto ad hoc, constituyendo una rama legítima de las matemáticas" (pág. 540). Lewin considera que, a pesar de

la falta de fundamentación matemática de su sistema, está en condiciones de continuar con el desarrollo de la teoría. De ahí su introducción, en el hipotético espacio hodológico, de construcciones que poseen un carácter direccional, como el concepto de "fuerza psicológica", que él mismo considera equivalente a lo que otros psicólogos denominan "impulso", "tendencia excitadora", etc.

Por citar un ejemplo bastante representativo del trabajo realizado en psicología en torno a la teoría de las catástrofes nos referiremos a los trabajos de Poston y Stewart (1978) y Stewart y Peregoy (1983) en el ámbito de la percepción. El punto de partida de este tipo de trabajos radica en el estudio de la percepción de figuras multiestables como, por ejemplo, el cubo de Necker. El fenómeno de percepción multiestable muestra que un estímulo único puede producir más de un efecto en el sistema perceptivo humano. Existen otros aspectos de la percepción que sugieren la posibilidad de un análisis en términos de estabilidad estructural. Por ejemplo, como señala Hochberg (1964), la percepción posee un carácter casi-contínuo, en el sentido de que pequeños cambios en el estímulo producen, generalmente, pequeños cambios en el percepto. Si consideramos el conjunto de los estímulos  $S$  y el conjunto de los perceptos  $P$  como dos espacios topológicos, entonces el producto  $S \times P$  dará lugar a una nueva topología.

El análisis de la percepción de figuras ambiguas (figura 5.3) sugiere algunas ideas interesantes en cuanto a la

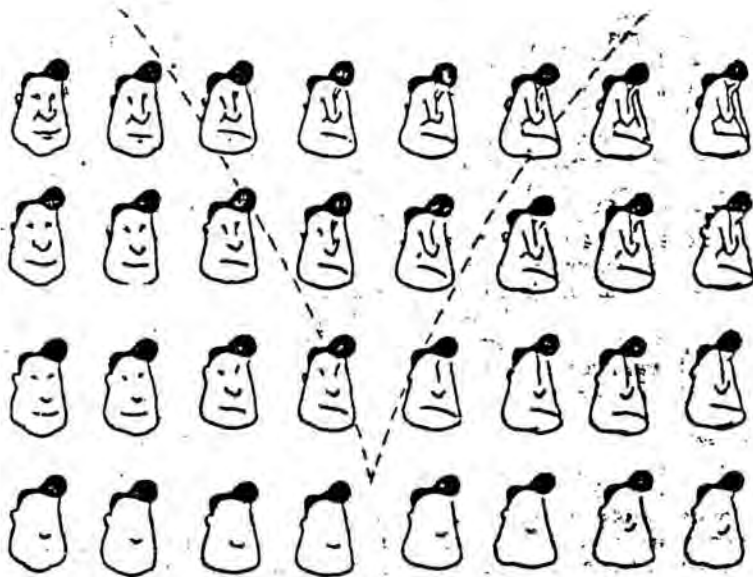


Figura 5.3 Figuras ambiguas utilizadas en el experimento de Stewart y Peregoy (1983)



modelización de procesos psicológicos basada en la teoría de las catástrofes. Este tipo de experimentos perceptivos suele basarse en la presentación secuencial de una serie de figuras ambiguas hasta que el sujeto experimental afirma que se ha producido un cambio cualitativo (por ejemplo, la figura ha dejado de representar a una mujer y pasa a representar un rostro masculino).

Como se ha indicado repetidamente, todo modelo basado en la TC (Teoría de las Catástrofes) contiene ciertas variables de control que supuestamente regulan la evolución del sistema. En este caso, la única variable de control considerada es el grado de ambigüedad de los estímulos presentados, aunque eventualmente podrían incluirse variables referidas al estado del sistema perceptivo, experiencia anterior, etc. De hecho, las nociones de variable control y variable independiente no coinciden completamente, puesto que una variable de control puede ser el resultado de la ponderación (e incluso una función no lineal) de varias variables independientes.

Dado que el fenómeno a estudiar es un cambio repentino en la calificación cualitativa (de hombre a mujer o viceversa), y dado también que se trabaja con una sola variable de control (el grado de ambigüedad de las figuras), puede plantearse la posibilidad de que el proceso subyacente sea descriptible en términos de la "catástrofe de pliegue" ("fold catastrophe"), una de las catástrofes elementales de Thom (ver apartado 5.1).

El pliegue, como las demás catástrofes elementales, está asociado con una función de energía  $V(x)$  que se supone que debe ser minimizada a lo largo del proceso. Esta idea de minimización de la función de energía puede relacionarse, al menos en principio, con algunas ideas formuladas en el campo de la percepción en cuanto a sus aspectos económicos (por ejemplo, el "principio de energía mínima"). La función asociada a la catástrofe de pliegue es, de acuerdo con la teoría de las catástrofes elementales,  $V(x) = X^3/3 + a x$ , siendo  $a$  la variable de control asociada a la variable independiente manipulada. Es necesario, a continuación, localizar los puntos críticos en los cuales el sistema está en equilibrio. A partir de la teoría de funciones, sabemos que dichos puntos son aquellos en los que la primera derivada de la función es igual a 0. Realizando esta operación, pueden hallarse todos los puntos que cumplen la condición de equilibrio (Figura 5.4). Para los valores negativos de  $a$  (comenzando la secuencia de presentación de estímulos por R1) la ecuación planteada al igualar a cero la primera derivada tiene dos soluciones: una de ellas (línea continua superior en la figura) es un mínimo local o atractor (dado que representa un mínimo de la función de energía); la otra (línea punteada superior), representa un máximo local y tiene un carácter de repulsión del proceso. Invirtiendo la secuencia de presentación de estímulos (comenzando por R8), puede realizarse la misma operación para valores positivos de  $a$  (líneas punteada y continua inferiores). Supondremos entonces que cuando la

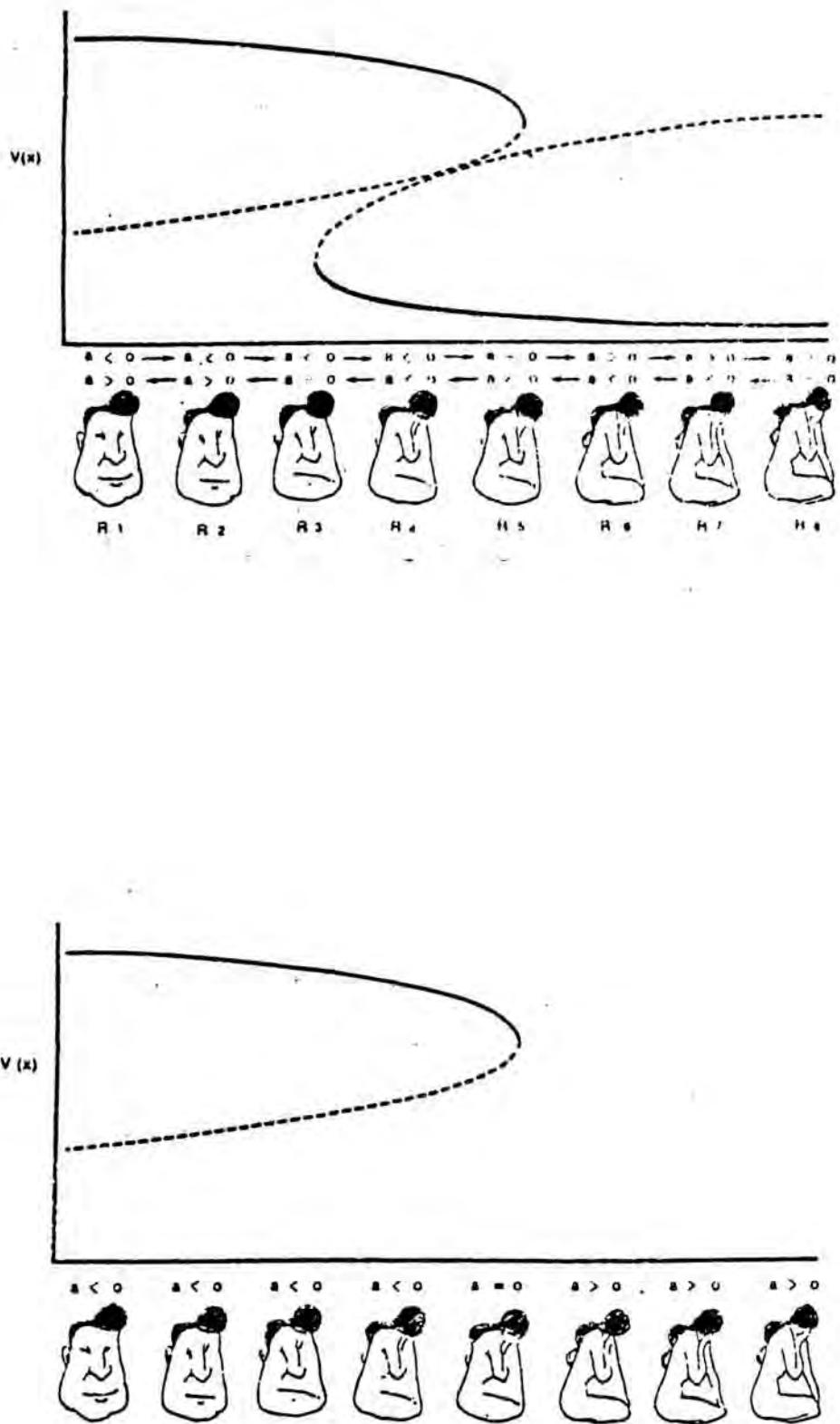


Figura 5.4 Gráficos para las catástrofes de cúspide (arriba) y de pliegue (abajo) en el experimento de Stewart y Peregoy (1983)

secuencia de estímulos llegue a  $a = 0$  se producirá un salto abrupto de una respuesta a otra. De acuerdo con el modelo, se produciría también un fenómeno de histéresis, puesto que el momento del "salto" de una respuesta a otra no coincide si la secuencia de estímulos comienza en  $R_1$  o en  $R_8$ . Resulta sugerente el hecho de que el modelo pueda dar cuenta del fenómeno de histéresis, puesto que éste se presenta con frecuencia en ámbitos como el de la psicofísica. De hecho, pensamos que, en general, los fenómenos tipo umbral pueden ser descritos muy adecuadamente por la teoría de las catástrofes elementales. Esta idea será también importante para cualquier intento de modelización matemática del funcionamiento de las unidades de un modelo distribucional interactivo que incluyan la presencia de un umbral de respuesta.

El hecho de que existan dos pliegues distintos y la constatación empírica del fenómeno de histéresis hacen pensar en la existencia de una dinámica compleja y en la necesidad de hallar otras variables controladoras del proceso. Stewart y Peregoy (1983) sugieren que una segunda variable importante en este contexto podría ser el grado de definición de los estímulos (con valores, por ejemplo, desde  $D_1$  hasta  $D_4$ , de mayor a menor definición).

Al margen del ejemplo concreto, debe señalarse que la introducción de una variable independiente complementaria sugiere la consideración de una segunda catástrofe elemental denominada "cúspide" o, más técnicamente, catástrofe de Riemann-Hugoniot. Aunque esta última denominación tiene su

origen en el ámbito de la hidronámica, actualmente es frecuente su utilización en campos de trabajo muy diversos. Por otra parte, la existencia de dos variables independientes complica la noción de variable de control, puesto que debemos definir ahora dos variables de control complejas que serán el resultado de la ponderación de las dos V.I. Uno de los objetivos de este tipo de modelización radica precisamente en evaluar la influencia de cada una de las variables independientes sobre las diversas variables de control que regulan el funcionamiento del sistema.

La figura 5.5 representa adecuadamente la posible evolución cualitativa del sistema de acuerdo con el modelo de cúspide. El plano de control representa las diversas combinaciones de los valores de las dos variables independientes consideradas. De acuerdo con el modelo de Riemann-Hugoniot, existirá una región multiestable que da lugar a bimodalidad, es decir, a la predicción de dos valores distintos de la variable estado (variable-efecto). En la proyección del plano de control puede observarse que dicha bimodalidad se representa geoméricamente como un pliegue en el plano representativo del comportamiento del sistema.

La existencia de una región multiestable debe interpretarse como sigue: en un modelo de cúspide podemos distinguir entre una variable "normal" que, para valores bajos de la otra variable de control, no produce cambios repentinos en la respuesta, y un factor de bifurcación, denominada así por el hecho de que aumentando su valor crece también la región en



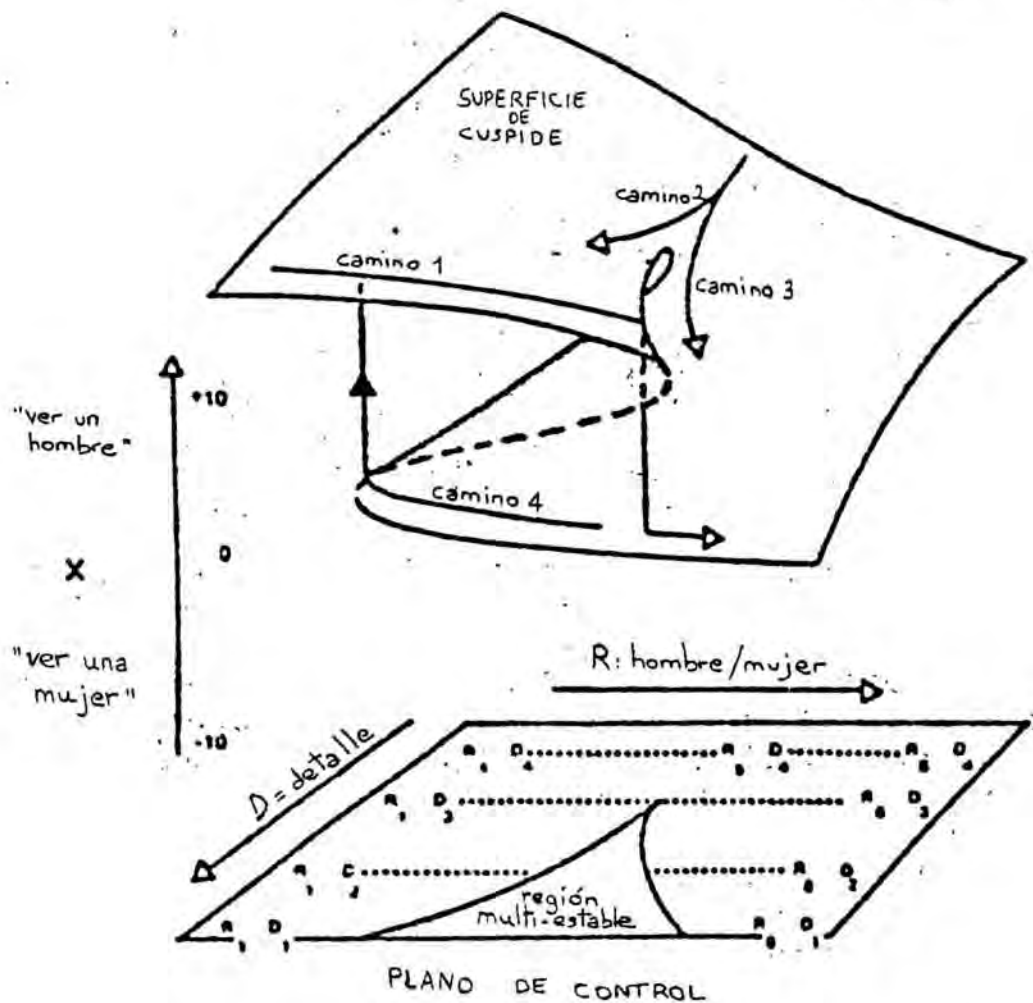


Figura 5.5 Proyección del plano de control en la catástrofe de cúspide

que se produce bimodalidad. En el caso que hemos estado comentando, la variable "definición de las figuras" posee un peso decisivo en el factor de bifurcación, de modo que para niveles bajos de detalle difícilmente se producirá un cambio en la respuesta del sujeto; sin embargo, a medida que aumenta el detalle de las figuras el rango de valores de la variable R en los que puede producirse un cambio "catastrófico" de respuesta se amplía progresivamente.

Este último punto plantea una cuestión de interés general, a saber, la posibilidad de derivar predicciones a partir de este tipo de análisis. Ciertamente, la ampliación de la región multiestable nos permite predecir la correspondiente ampliación del rango de valores de R en los que puede producirse el cambio de respuesta. En otras palabras, la dispersión de los puntos de cambio de respuesta será mayor para valores altos de detalle de las figuras que para valores bajos de esa misma variable. Sin embargo, queda pendiente un tema básico: ¿es posible derivar predicciones más precisas, es decir, es posible extraer consecuencias cuantitativamente precisas de este tipo de planteamiento?. En principio, muchos críticos de la teoría (y algunos de sus defensores) responderán negativamente, arguyendo su carácter fundamentalmente cualitativo. De acuerdo con estos críticos, la teoría de las catástrofes se limitaría a presentar un conjunto de "recetas" destinadas a atribuir a un conjunto de datos un determinado modelo cualitativo (por ejemplo, un modelo de cúspide). El propio Thom (1980/1985) afirma explícitamente

que su intención no es la de derivar predicciones cuantitativas. De acuerdo con su punto de vista, si se desea modelizar cuantitativamente situaciones locales hay que introducir un número suficiente de parámetros; pero si lo importante es aislar las grandes regularidades subyacentes a los fenómenos, entonces será fundamental contentarse con un número pequeño de parámetros

representativos, lo cual implicará mantenerse lejos de la cuantificación extrema propia de algunos modelos.

A pesar de lo expuesto, pensamos que es posible una aproximación cuantitativa desde la teoría de las catástrofes. Siguiendo a Poston y Stewart (1978) y a Stewart y Peregoy (1983), nos parece fundamental señalar la diferencia entre dos aspectos básicos: Por una parte, la teoría permite realizar una clasificación cualitativa de cierta nube de puntos, asignándole hipotéticamente alguno de los modelos de la teoría de las catástrofes elementales. Sin embargo, ¿por qué debe detenerse el proceso en este punto?; en otras palabras, si se determina que cierto proceso es descrito por un modelo de cúspide, ¿no será posible determinar qué cúspide es la que mejor se ajusta a los datos obtenidos?. La respuesta nos parece afirmativa, sobretodo a partir de los estudios presentados por Cobb (1978, 1980) en cuanto a la utilización del método de máxima verosimilitud para la determinación de los valores de los parámetros del modelo.

Aunque no entraremos en consideraciones excesivamente técnicas, sí nos parece interesante indicar en qué puntos básicos

radican las dificultades de la teoría de las catástrofes en lo que respecta al análisis cuantitativo. La dificultad básica se encuentra, obviamente, en los principios topológicos que fundamentan la teoría. El problema es que las medidas de bondad de ajuste existentes no son invariantes bajo las transformaciones permisibles en el marco de la topología diferencial. Por ejemplo, el criterio tradicional de mínimos cuadrados se mantiene invariante únicamente bajo transformaciones lineales (por ejemplo, cambios lineales en la escala del sistema de coordenadas). Sin embargo, el criterio de equivalencia topológica se basa en la existencia, entre dos sistemas de coordenadas, de aplicaciones infinitamente diferenciables, es decir, de difeomorfismos (que implican cambios de coordenadas curvilíneas).

Otro aspecto "problemático" de la teoría de las catástrofes radica en el hecho de que una misma configuración de variables de control puede existir más de un valor estable de la variable estado. Esto plantea una dificultad suplementaria, puesto que casi todas las técnicas de estimación de parámetros se basan en la minimización de las diferencias entre los datos empíricos y las predicciones del modelo. Sin embargo, este problema pierde sentido si se dispone de datos en forma de series temporales, puesto que en este caso puede considerarse como variable dependiente el cambio en la variable conducta; este cambio es un valor único de una función polinómica de las variables independientes. Si se asume que los datos empíricos son una "muestra instantánea" de una

población de sistemas dinámicos estocásticos, entonces para cada configuración de las variables de control existirá una distribución de frecuencias multimodal de la variable conducta.

A partir de un análisis de los modelos matemáticos del proceso de difusión, Cobb (1978) demuestra que la función de densidad de probabilidad que da lugar a las distribuciones empíricas del tipo que estamos considerando es miembro de una familia canónica de funciones de densidad de tipo exponencial. La ventaja fundamental de este tipo de funciones es que para cada una de sus "familias" pueden definirse un conjunto de estadísticos y, además, existen estimadores de máxima verosimilitud para los parámetros de cada familia. Existía, sin embargo, la dificultad práctica de que los sistemas utilizados para realizar las estimaciones eran extremadamente lentos, puesto que implicaban la resolución de integrales en cada una de las iteraciones. Cobb (1978) presenta un nuevo procedimiento de estimación que exige únicamente la solución de un sistema de ecuaciones lineales. En un trabajo posterior (Cobb, 1980), se presenta un programa de ordenador capaz de realizar automáticamente la estimación requerida para modelos de cúspide.



### 5.3 Hacia la construcción de modelos del cambio en psicología

Presentaremos en este apartado un conjunto de ideas dirigidas a la modelización tanto de los aspectos regulares como de las discontinuidades que, de acuerdo con nuestro punto de vista, se presentan en el funcionamiento del sistema cognitivo humano. Algunas de las propuestas permiten un análisis matemático bastante preciso. Otras, en cambio, se mantiene todavía en un nivel meramente intuitivo, aunque, en nuestra opinión, existen amplias posibilidades para su desarrollo futuro.

Una idea fundamental para este trabajo la constituye la consideración del sistema cognitivo humano (o, al menos, de una parte substancial del mismo) como análogo a un sistema gradiente. Intentaremos mostrar inmediatamente, partiendo de un ejemplo bastante general, cómo las formulaciones más avanzadas de los modelos distribucionales interactivos tienden a converger en el punto de vista que estamos sosteniendo. Al propio tiempo, realizaremos una primera aproximación al tipo de fenómenos que, a nuestro entender, deben ser aprehendidos en una modelización adecuada del sistema cognitivo humano.

Consideremos, por ejemplo, el proceso de solución de problemas en sujetos humanos. Como se indicaba ya en el capítulo anterior, existen razones importantes para suponer que la solución de cierto tipo de problemas es el resultado de cambios cualitativamente importantes en la dinámica "microscópica" subyacente. En el ejemplo que mencionábamos en el capítulo anterior, ese cambio cualitativo radicaba en el salto de un dominio de solución a otro como consecuencia de la

ruptura de la fijación perceptiva inducida por el material experimental. En el marco de los modelos distribucionales interactivos que estamos defendiendo, esto se traduciría, lógicamente, en modificaciones en los estados de equilibrio del sistema de unidades elementales. Supondremos que el funcionamiento de este sistema presenta, a lo largo del proceso de solución (y en cualquier otra situación dada) un conjunto de fluctuaciones de carácter esencialmente aleatorio. En condiciones normales, estas fluctuaciones no se amplificarán y tenderán, en consecuencia, a "regresar" a la situación de equilibrio anterior.

De acuerdo con los trabajos clásicos de la psicología de la Gestalt (Köhler, 1947; Wertheimer, 1945), consideraremos que la presentación de una situación-problema producirá un aumento en la "tensión" del sistema, cambio que interpretaremos como un alejamiento de la situación de equilibrio, la cual, a su vez, será definida como aquella que presente un nivel global de "energía" mínimo. La energía del sistema puede ser definida, como veremos algo más adelante, en función de la relación entre el estado del sistema cognitivo y su entorno; como veremos, éste introduce, por definición, una serie de restricciones a las cuales el sistema debe acomodarse (por ejemplo, las características de la solución o las "reglas del juego" en una situación-problema). Sostenemos también que cuando el nivel de "tensión" supera un determinado umbral las fluctuaciones aleatorias en el funcionamiento del sistema pueden progresar y ampliarse. En

este último caso, el resultado será una bifurcación en el funcionamiento del sistema, que a su vez se traducirá en cambios cualitativos como los que se producen, por ejemplo, en el salto repentino de un dominio de solución a otro, o en la experiencia subjetiva descrita por los psicólogos de la Gestalt como "comprensión súbita".

La problemática que estamos analizando presenta algunos puntos de contacto con la que se plantea en ciertos ámbitos de la física. El alejamiento de la situación de equilibrio en sistemas donde interactúa un gran número de unidades es producto de un conjunto de ligaduras externas cuyos valores obligan al sistema a alcanzar un estado lejano al de equilibrio. Otra característica básica de este tipo de sistemas, como indicábamos en el capítulo 3, es la existencia de ciertos mecanismos de interacción no lineal entre los elementos del sistema. En estas condiciones, puede producirse, al contrario que en las estructuras estables, un comportamiento cooperativo de un gran número de unidades que se traduzca en la consecución de un estado de equilibrio y de un conjunto de relaciones entre elementos bastante distintas de las que se producían anteriormente.

El origen de las fluctuaciones locales en un sistema distribucional bajo condiciones de baja tensión puede encontrarse en cambios aleatorios poco importantes en la activación de un conjunto de unidades. De hecho, dado el carácter fuertemente interactivo de este tipo de sistemas, estos cambios pueden deberse a la influencia del conjunto del sistema

sobre cada una de las unidades de acción, sin que pueda descartarse tampoco la presencia de cambios aleatorios espontáneos en la activación de unidades debidos a razones "internas" de las propias unidades. La consecuencia de la superposición de estas pequeñas perturbaciones es que las ecuaciones que describen el cambio presentan una forma inicialmente lineal.

Consideramos, que este tipo de fluctuaciones, amplificadas bajo condiciones de alta tensión, pueden ser parcialmente responsables, por ejemplo, de que en el problema de los 9 puntos el sujeto alargue repentinamente las líneas de solución más allá del marco del cuadrado supuestamente definido por esos puntos. En términos poco precisos, este fenómeno sería el resultado de la activación de un conjunto de unidades que representarían, por ejemplo, la imagen de una línea que supera el marco de los 9 puntos. Si atribuimos a esta fluctuación un carácter puramente aleatorio, deberemos considerarla como una distorsión en el funcionamiento de las unidades implicadas en la tarea. Sin embargo, no nos parece improbable la presencia de aspectos más direccionales en este tipo de proceso. En todo caso, lo que nos parece fundamental es el cambio en la probabilidad de amplificación de la fluctuación en función del estado tensional del sistema. Algunas líneas más atrás hemos presentado un principio de carácter general que enunciaremos ahora de modo preciso: el sistema de unidades de tipo distribucional mantiene en todo momento un conjunto de valores de activación de unidades y

de "pesos" de sus interacciones tal que la energía total del sistema sea la mínima posible, dadas las condiciones del entorno. De esta forma, la consecuencia del aislamiento del sistema sería necesariamente su evolución hasta un estado de equilibrio que, como señalábamos en el capítulo anterior, es comparable con su "muerte termodinámica".

Consideremos, en primer lugar, un modelo distribucional que incluya unidades cuyo funcionamiento está asociado a un umbral de activación. De acuerdo con Ackley, Hinton y Sejnowski (1985), la energía correspondiente a una configuración global de activación cualquiera vendría expresada en la siguiente ecuación:

$$E = -\sum_{i < j} W_{ij} S_i S_j + \sum_i U_i S_i$$

donde  $W_{ij}$  es la fuerza o "peso" de la conexión entre unidades  $i$  y  $j$ ;  $S_i$  es 1 si la unidad  $i$  está en situación de activación, y 0 en caso contrario.  $U_i$  es el umbral de activación para la unidad  $i$ .

Supongamos que el sistema está ejecutando cualquier tarea (por ejemplo, interpretar una imagen, o solucionar un problema). Será posible, entonces, referirse a la diferencia de energía entre dos configuraciones: la que resultará de aceptar una determinada hipótesis (sobre la interpretación de la imagen, o sobre el camino de solución de un problema), y la que resultaría de no aceptarla. En términos más generales, podremos referirnos a la diferencia de energía entre los patrones de activación asociados a dos estados mentales distintos. Siguiendo con unidades "tipo umbral", la diferencia



de energía entre las dos configuraciones vendrá dada por la expresión:

$$\Delta E_k = \sum_i W_{ki} S_i - U_k$$

Entonces, la regla para minimizar la energía aportada por una unidad concreta consistirá en adoptar el estado activado si el input total de las unidades conexas supera el umbral. Debe señalarse que, aunque este tipo de regla de minimización es típica de las unidades que presentan umbrales de activación, puede ser generalizada fácilmente a otro tipo de unidades. En un elemento de tipo absolutamente continuo (es decir, sin umbral de activación), la regla de minimización consistiría en aumentar o disminuir el estado de activación de forma proporcional al input recibido. Este cambio de tipo proporcional es, como veíamos en el capítulo anterior, un proceso fundamental para muchos modelos distribucionales. Debemos remarcar, además, algo que señalábamos también en el capítulo 4: el hecho de introducir un umbral no convierte automáticamente un elemento en una unidad tipo puerta lógica; puede conservarse la noción de niveles continuos de activación por encima y por debajo del umbral, de forma que un mismo input procedente de las unidades conexas puede producir o no la activación de la unidad, según su nivel anterior estuviera más o menos próximo al umbral. Se trata, de hecho, de algo muy parecido al umbral diferencial en psicofísica. Hopfield (1982) sugiere la existencia, en este tipo de redes de unidades, de mínimos locales y globales de energía, y propone algunas ideas relacionadas con la transición del

proceso a lo largo de los diversos niveles energéticos. En especial, Hopfield distingue entre algunas situaciones en las que la evolución más adecuada del sistema es la que le dirige hacia un mínimo energético local; en otros contextos (y, particularmente, ante restricciones importantes procedentes del entorno), será más adecuada la transición hacia un mínimo global. Debe señalarse inmediatamente que los dos tipos de transición que acabamos de mencionar se derivan del análisis de los sistemas gradientes y corresponden a lo que, en el marco de la teoría de las catástrofes, hemos denominado "regla de Maxwell" (transición hacia el mínimo global) y "regla de retraso perfecto" ("delay rule", transición hacia un mínimo local).

Metropolis, Rosenbluth, Rosenbluth, Teller y Teller (1953) propusieron, en el marco del estudio de las propiedades de sistemas termodinámicos, un algoritmo que posee la propiedad de evitar los mínimos locales de energía para dirigir la transición hacia el mínimo global. Este tipo de planteamiento ha sido aplicado mucho más recientemente al problema de la acomodación a las restricciones del medio (Kirkpatrick, Gelatt y Vecchi, 1983). Ackley, Hinton y Sejnowski (1985) retoman esta idea en una forma adecuada a los sistemas estructurados en paralelo que ejecutan operaciones de procesamiento. Si la diferencia energética entre los estados "activado" o "desactivado" de la unidad  $k$  es  $E_k$ , entonces la probabilidad de que la unidad  $k$  supere el umbral de activación es igual a:

$$P_k = 1 / (1 + e^{-\Delta E_k / T})$$

donde T es un parámetro que actúa con la misma función que la temperatura. Una red de unidades que sigan esta "regla de decisión" puede alcanzar eventualmente un estado de equilibrio "termal"; la probabilidad relativa de dos estados energéticos globales seguirá la distribución de Boltzmann, es decir:

$$P_a / P_b = e^{-(E_a - E_b) / T}$$

donde  $P_a$  es la probabilidad de que el sistema derive hacia el estado A, y  $E_a$  es la energía asociada a ese estado.

La distribución de Boltzmann se relaciona muy estrechamente con la teoría de la información. En particular, la diferencia entre los logaritmos de las probabilidades de dos estados globales es precisamente su diferencia de energía para un valor de T igual a 1.

La formulación de Hopfield (1982) y de Ackley, Hinton y Sejnowski (1985) corresponde muy exactamente a lo que, en el campo de la teoría de las catástrofes (y en ámbitos mucho más generales) es denominado sistema gradiente. En términos thomianos, se parte de la definición de un abierto W en el espacio-tiempo y, sobre este abierto, de un campo de dinámicas locales (M, X). La variedad M es el espacio de los estados internos del sistema; todo punto x de W tiene asociado un campo de vectores X(x) de M que dependen diferenciablemente de las coordenadas de x. El campo X(x) es, para todo punto x de W, un campo gradiente sobre M, de forma que:

$$X(m; x) = - \text{grad } V(m; x)$$

donde  $V(m;x)$  es una función potencial sobre  $M$  que depende diferenciablemente de  $x$ .

En un punto  $x$  de  $W$ , los regímenes locales estables son definidos por los atractores estructuralmente estables del campo  $X$ . Si aceptamos algún tipo de principio económico, estos atractores corresponderán a los valores mínimos del potencial  $V(m;x)$ . En general, además de un mínimo global, existirá también un conjunto de mínimos relativos de carácter local. En este caso, como señalan Isnard y Zeeman (1976) la utilización de cualquiera de las dos convenciones posibles deberá ser justificada en cada caso concreto.

La formulación de Ackley, Hinton y Sejnowski (1985) toma explícitamente el nombre de "máquina de Boltzmann". Como señalábamos en el capítulo anterior, la denominación resulta muy adecuada y sugerente. Pero, como señalábamos casi al mismo tiempo, esta formulación hereda algunos de los problemas del enfoque boltzmanniano y, muy particularmente, su carácter "excesivamente regular" o continuo. El problema es el mismo que se plantea a la termodinámica clásica, ligado esencialmente al carácter necesariamente creciente de la magnitud entropía. A partir del enfoque estadístico de Boltzmann, la entropía contendrá el "grado de desinformación" del observador macroscópico y representará el número de configuraciones microscópicas compatibles con cierto estado de equilibrio. La entropía de Boltzmann asocia al sistema una idea de orden, de manera que el segundo principio de la

termodinámica clásica se convierte en una ley de desorganización progresiva. Como indicábamos en el capítulo 3, este tipo de formulación es parte fundamental de una cierta visión de la ciencia que parece arrinconar los fenómenos vitales (y, por tanto, los psicológicos) al ámbito de lo "radicalmente extraño". Siguiendo en la línea que exponíamos en el mencionado capítulo, afirmamos explícitamente que la formulación de un sistema distribucional interactivo para los procesos cognitivos en términos meramente boltzmannianos no puede dar cuenta de todo un conjunto de cambios cualitativos que, a nuestro entender, forman parte inextricable de los procesos psicológicos. De ahí, por una parte, la formulación del concepto de estructura disipativa y, por otra, la necesidad, que nosotros afirmamos, de recurrir a una visión semejante a la de la teoría de las catástrofes para la modelización de lo que hemos denominado, en un sentido muy general, el cambio. Intentaremos avanzar en esta línea proponiendo un ejemplo de utilización de la teoría de las catástrofes en el campo de la solución de problemas. Nuestra pretensión es más la de ilustrar un planteamiento que hemos venido exponiendo repetidamente que la de formular un modelo ideal o exacto en todos sus detalles (aunque, naturalmente, intentaremos aproximarnos en lo posible a este segundo objetivo).

A nuestro entender, debemos aportar inicialmente una interpretación no puramente matemática del concepto de potencial. Creemos haber demostrado que la utilización de conceptos



energéticos es perfectamente coherente en el marco de los modelos distribucionales que venimos defendiendo. Sin embargo, se hace necesaria una interpretación adecuada de este tipo de conceptos en el marco de la psicología cognitiva y, especialmente, en el ámbito que estamos considerando últimamente (solución de problemas). En nuestra opinión, esta interpretación puede formularse en términos de **distancia a la solución**. El concepto gestaltista de tensión producida por la presentación de un problema y el de potencial de un sistema gradiente adquieren mayor sentido si consideramos que el sistema compara, de algún modo, su estado actual con un estado ideal final que correspondería a la representación de la solución del problema. Esta representación puede tomar diversas formas, dependiendo de la naturaleza del problema y de la propia solución: una imagen (por ejemplo, la de una torre de Hanoi correctamente construida), una descripción verbal, o una combinación de ambas. En cualquier caso, y atendiendo al carácter unificado de la dinámica subyacente a tales representaciones (y, en particular, a la equivalencia topológica que les atribuimos), parece factible afirmar la posibilidad de tales procesos comparativos, que podrían ser ejecutados por módulos especializados, de forma que los pesos conexionales entre las distintas unidades de tales módulos reflejaran el grado de disimilaridad entre el estado de las unidades que representan la situación actual y el estado de las unidades que representarían la situación final (solución del problema). Este ajuste continuo de los pesos de

tales módulos puede ser realizado en la forma habitual, mediante aplicación de la regla Delta. Además, deberemos suponer que los módulos destinados a tal función poseerán un rico conjunto de interconexiones con diversas regiones del sistema, de forma que los cambios en su activación puedan ser responsables de modificaciones de tipo relativamente global.

Consideremos ahora una tarea experimental concreta. Para generalizar un poco más el planteamiento, nos centraremos inicialmente en un problema distinto al que hemos comentado algunos párrafos atrás, a saber, en un problema de jarras de agua. Como es sabido, en este tipo de problema el sujeto debe intentar medir una cierta cantidad de líquido realizando una serie de trasvases entre recipientes, ninguno de los cuales, obviamente, permite medir directamente la cantidad requerida. Intentaremos construir un modelo del fenómeno que Scheerer (1963) denominó "fijación por hábito". El fenómeno consiste en que la presentación de un conjunto de problemas de entrenamiento que poseen la misma solución induce a los sujetos a seguir utilizando esa solución en los problemas propiamente experimentales, aun en el caso de que éstos admitan una solución más sencilla. Resulta claro que, en estas circunstancias, pueden presentarse (y se dan de hecho) conjuntos de respuestas de carácter bimodal. A fin de mejorar la precisión del análisis, utilizaríamos como elemento comparativo el árbol de soluciones de cada uno de los problemas

y valoraríamos la proporción de sujetos que utiliza cada uno de los posibles caminos de solución. Lógicamente, a mayor dificultad del problema (por ejemplo en cuanto a número de recipientes), corresponderá una mayor complejidad del árbol de soluciones y, en consecuencia, un mayor grado de precisión en el análisis.

Como se ha indicado en el apartado anterior, los factores de control propios de un modelo basado en la teoría de las catástrofes no deben coincidir necesariamente con las variables independientes implicadas en una tarea experimental. En general, se asume que los distintos factores de control son el resultado de una ponderación lineal entre las distintas variables independientes aunque, naturalmente, éstas pueden poseer pesos muy distintos en los diversos factores de control. Sin embargo, en aras de la simplicidad, intentaremos inicialmente identificar las distintas variables independientes propias de este tipo de tareas con los distintos factores de control. Por supuesto, sólo la posible contrastación del modelo podría apoyar en mayor o menor medida estas presuposiciones.

De acuerdo con lo expuesto, consideraremos el número de problemas de entrenamiento como lo que, en el marco de la teoría de las catástrofes, se denomina "factor normal". Si consideráramos únicamente esta variable, podría presuponerse lógicamente la presencia de bimodalidad para los valores centrales del factor, pero no para los valores extremos.

Es factible introducir inmediatamente un factor de bifurcación importante, que se relacionaría con una variable que

podemos definir como "diferencia de dificultad entre las dos soluciones del problema". Como se ha indicado, los problemas experimentales presentan dos soluciones posibles, de las cuales la más difícil ha sido "fijada" a través de la presentación de los problemas de entrenamiento. Podemos considerar que las soluciones "difíciles" para los distintos problemas presentan un nivel de complicación semejante. De este modo, la diferencia entre las dos soluciones dependerá únicamente de la dificultad que presente la solución "fácil". Lógicamente, la región multiestable definida por el factor de bifurcación sería más amplia para valores grandes de la diferencia; en cambio, cuando la solución "fácil" resulta ser tanto o más complicada que la solución "difícil", la región multiestable probablemente se reduzca, puesto que los sujetos difícilmente se apercibirían de la existencia de una solución alternativa. Sin embargo, ello puede depender en buena medida del rango de valores considerado. Por otra parte, esta variable podría tener también un efecto de "sesgo" sobre el modelo de cúspide que acabamos de definir, por lo que será útil introducir un tercer factor que pueda incorporar este aspecto.

La introducción de un factor de "sesgo" nos permitirá también introducir en el modelo una nueva variable. Consideramos importante valorar el efecto del tiempo transcurrido entre la realización de los problemas de entrenamiento y la presentación de los problemas experimentales. Resulta claro que el efecto del aumento de valor de esta variable debería

ser el de sesgar el conjunto de las respuestas (independientemente de los demás factores) en favor de las soluciones más fáciles (o, más exactamente, disminuir la fijación en las soluciones "difíciles"). Sin embargo, como ocurría en parte con la diferencia de dificultad entre las dos soluciones, la variable que estamos comentando parece presentar una multiplicidad de efectos. De hecho, podría pensarse también en un efecto de bifurcación, al menos para intervalos relativamente largos. Naturalmente, esto no representa ninguna dificultad para el modelo puesto que, como hemos venido indicando, cada una de las variables independientes implicadas puede relacionarse con los diversos factores de control; de esta forma, puede aprehenderse simultáneamente el efecto de sesgo de una variable y su posible efecto bifurcativo.

Podría valorarse también la posibilidad de introducir alguna variable relacionada con aspectos motivacionales (por ejemplo, la introducción de algún tipo de incentivo destinado a primar un tipo determinado de soluciones). Probablemente esta variable se relacionaría primordialmente con el factor de sesgo. En cualquier caso, resulta claro que este tipo de modelos permite, a partir de la noción básica de factor de control como ponderación de diversas variables independientes, introducir un buen número de tales variables sin aumentar excesivamente la complejidad del modelo.

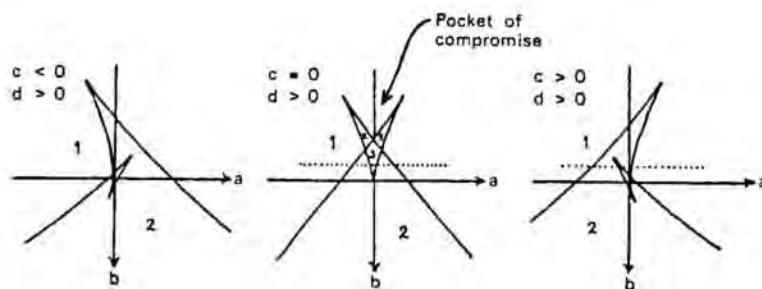
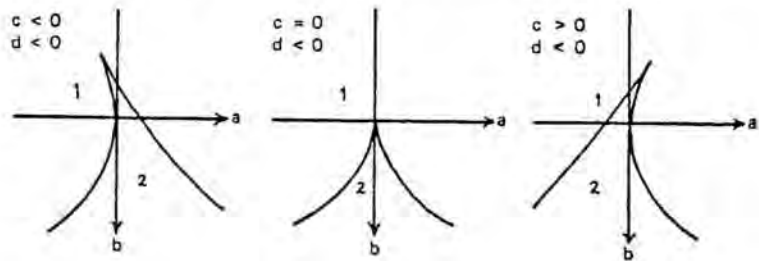
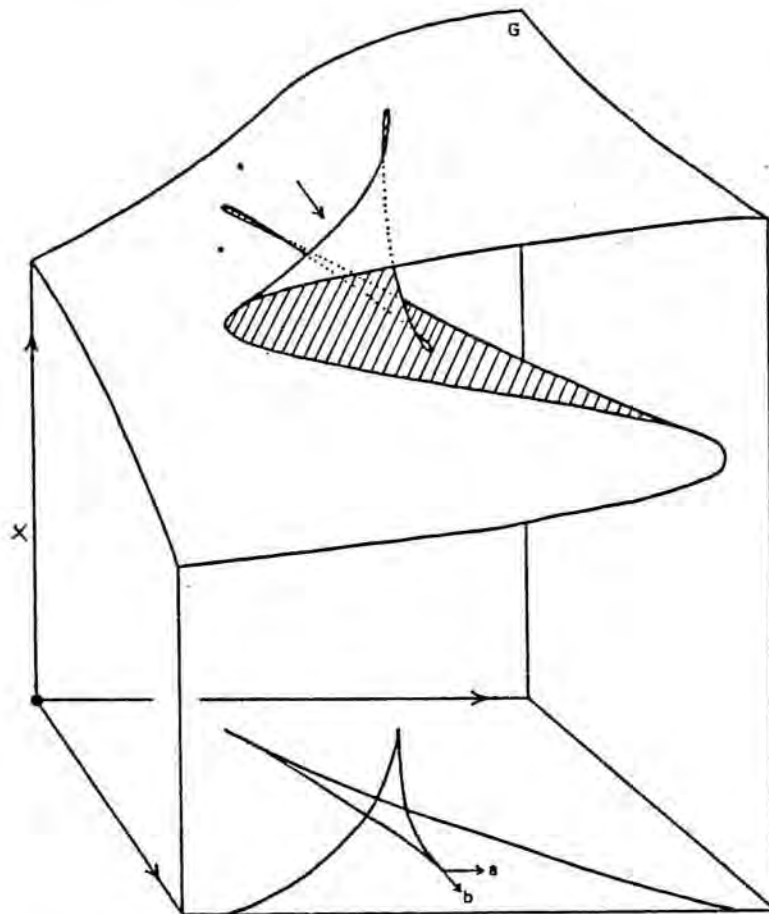
Si no añadimos ningún otro elemento en el modelo, éste nos quedará definido por tres factores de control (factor normal, factor bifurcativo y factor de sesgo), de manera que



tomará la forma de la catástrofe elemental de "cola de golondrina". Como es sabido, la expresión que relaciona el potencial del sistema con sus variables de control y de estado es, en el caso de la catástrofe que estamos considerando, la siguiente:  $x^5/5 + ux^3/3 + vx^2/2 + wx$ . Naturalmente, el paso subsiguiente consistirá en hallar la primera derivada de esta expresión a fin de encontrar los mínimos y máximos de la función. En el caso de la "cola de golondrina" podrán encontrarse dos mínimos de la función de potencial, que corresponderán a los dos atractores del sistema: el correspondiente a la solución "fácil" y el que correspondería a la solución "difícil" del problema.

Podría valorarse la posibilidad de introducir un cuarto factor que jugara el papel de lo que, en el marco de la teoría, se ha denominado "butterfly factor" (factor-mariposa). El nombre de este factor es debido a que su introducción da lugar a una nueva catástrofe elemental, algo más compleja que la anterior, y en la cual la proyección del plano de control adopta una forma algo parecida a una mariposa (figura 5.6). Esencialmente, el papel del factor mariposa suele ser el de bifurcar la cúspide obtenida por introducción de los dos primeros factores, mientras que el factor de sesgo seguiría ejerciendo su propio efecto de forma relativamente independiente. El resultado de tal bifurcación es, para valores positivos del factor-mariposa, la emergencia de un tercer atractor que se sitúa en un valor intermedio entre los dos mínimos de la función de potencial que comentábamos

figura 5.6 Proyección y secciones de la catástrofe de "mariposa"



en el caso de la catástrofe de "cola de golondrina". Dado su carácter central, este atractor suele asociarse a la idea de "compromiso" entre los dos atractores anteriores. La idea es entonces que, a medida que el factor-mariposa aumenta de valor, mayor es la probabilidad de que el proceso derive hacia él.

Pensamos que sería interesante valorar la posibilidad de que el tiempo juegue en nuestro modelo el rol de factor-mariposa. La idea subyacente sería, como insinuábamos ya algunos párrafos más atrás, que el transcurso del tiempo jugaría un papel de incremento en la tensión global del sistema, obligando así a la adopción de un criterio de compromiso. Este criterio evitaría la prolongación excesiva de la situación-problema y limitaría, por tanto, el valor máximo de tensión al cual está sometido el sistema.

Aunque la idea que acabamos de exponer sobre el efecto del tiempo nos parece bastante verosímil, en el caso concreto que estamos examinando choca con un problema evidente referido al tipo de tarea que estamos analizando. En principio, los problemas que hemos planteado presentan un árbol de soluciones bastante limitado, con lo cual la referencia a "soluciones de compromiso" puede resultar un tanto forzada. Sin embargo, este hecho no elimina la posibilidad (ni el interés) de que en problemas en los que este tipo de solución pueda plantearse de modo más natural pueda utilizarse un modelo que incluya un factor-mariposa. Pensemos, por ejemplo, en un problema de ajedrez. Al margen de los tres

factores inicialmente introducidos (que, naturalmente, corresponderán en esta tarea a variables muy distintas), el factor tiempo puede jugar el papel de factor-mariposa cuando, bajo circunstancias de limitación de tiempo (muy frecuentes en ajedrez) el sujeto no esté en condiciones de finalizar el proceso de evaluación de movimientos posibles con la profundidad deseada. O, en otro caso, imaginemos que se produce la típica dicotomía entre la posibilidad de mejora de la posición que introduce un cierto movimiento y el riesgo asociado al mismo, o la divergencia entre mejora de posición y pérdida de material. En todos estos casos, el sujeto se enfrenta a una situación en la que deba decidir entre soluciones que presenten características bastante distintas. Resulta bastante natural, entonces, afirmar que bajo condiciones de premura de tiempo los sujetos pueden tender hacia una opción de compromiso. Si esta idea se ve apoyada por un conjunto apreciable de datos empíricos (y existen indicios de que es así; Newell y Simon, 1972), entonces la utilización de un modelo de catástrofe elemental de mariposa parece muy adecuada.

A partir de estas consideraciones iniciales, el proceso a seguir sería idéntico al que exponíamos en el caso de la catástrofe de "cola de golondrina". Para la "catástrofe elemental de mariposa" la expresión de la función de potencial es la siguiente:  $\frac{x^6}{6} + tx\frac{4}{4} + ux\frac{3}{3} + vx\frac{2}{2} + wx$ . Tras igualar la primera derivada a 0 puede demostrarse la presencia de tres mínimos de la función de potencial los cuales, una vez más,

corresponderían a los tres atractores del sistema (introduciéndose, a diferencia del caso anterior, el "atractor de compromiso"). El problema radicaría entonces, una vez establecidas las características cualitativas de la dinámica que estamos considerando, en intentar una aproximación cuantitativa semejante a la propuesta por Cobb (1980) para las catástrofes de cúspide. La aplicación de este tipo de análisis a catástrofes más complejas presenta indudables complicaciones y se encuentra solamente en sus inicios. Resulta evidente, en todo caso, que el desarrollo de este tipo de técnicas debe formar parte de un programa general destinado a potenciar la utilización de la teoría de las catástrofes (o, más en general, de teorías de la estabilidad estructural) en el ámbito de la psicología.



## A MODO DE CONCLUSION

No insistiremos una vez más sobre los objetivos que planteaba esta Tesis. A lo largo del trabajo hemos intentado extraer conclusiones claras respecto que a los distintos aspectos de la problemática que hemos analizado, y al propio trabajo nos remitimos para su valoración. En términos globales, nuestra opinión es que los modelos distribucionales interactivos constituyen una alternativa fundamentada y muy prometedora ante las dificultades casi insalvables que asedian al enfoque computacional-discreto. Creemos, no obstante, que pasar de la máquina de Turing a la máquina de Boltzmann no es suficiente. Después de todo, y si se nos perdona la redundancia, la máquina de Boltzmann no deja de ser una máquina. Sin embargo, si atribuimos a esa máquina un conjunto de propiedades de tipo global quizá podamos enfrentarnos a un conjunto de fenómenos cuyo carácter holístico y en muchos casos aparentemente rebelde a la formalización les ha mantenido alejados de la corriente principal de investigación en psicología cognitiva.

Creemos que la formulación del tipo de modelos que hemos propuesto significa la construcción (o, al menos, el proyecto de construcción) de una conexión entre el desarrollo reciente de ciertos aspectos de las matemáticas y de las ciencias fácticas más elaboradas y el ámbito de la psicología cognitiva. Hemos intentado dar los primeros pasos para cruzar ese puente. La clásica pregunta de la teoría de

las catástrofes ("el puente, ¿resistirá o se hundirá?") queda por responder, pero al menos podemos garantizar que sus fundamentos parecen bastante sólidos. Pensamos, en todo caso, que intentar cruzar ese puente es una empresa lo bastante atractiva como para asumir el riesgo de un evento catastrófico. Después de todo, este tipo de sucesos parece encontrarse en la misma esencia del cambio.

## REFERENCIAS BIBLIOGRAFICAS

- Ackley, D.H., Hinton, G.E. y Sejnowski, T.J. (1985). A learning algorithm for Boltzmann Machines. **Cognitive Science**, 9, 147-169.
- Adams, M.J. (1979). Models of word recognition. **Cognitive Psychology**, 11, 133-176.
- Allport, D.A. (1979). Conscious and unconscious cognition: A computational metaphor for the mechanism of attention and integration. En L.G. Nilsson: **Perspectives in Memory Research**. Hillsdale: Lawrence Erlbaum.
- Allport, D.A. (1980). Patterns and actions. En G. Claxton: **Cognitive Psychology: New directions**. London: Routledge.
- Allport, D.A. (1984). Alternatives to the computational view of the mind: The baby and the bathwater. **Journal of Verbal Learning and Verbal Behavior**, 23, 315-324.
- Anderson, J.A. & Hinton, G.E. (1981). Models of information processing in the brain. En G.E. Hinton y J.A. Anderson: **Parallel Models of Associative Memory**. Hillsdale: Erlbaum
- Anderson, J.A. y Mozer, M.C. (1981). Categorization and selective neurons. En G.E. Hinton y J.A. Anderson: **Parallel models of associative memory**. Hillsdale: Erlbaum
- Anderson, J.A., Silverstein, J.W., Ritz, S.A. & Jones, R.S. (1977). Distinctive features, categorical perception and probability learning. Some applications of a neural model. **Psychological Review**, 84, 413-451.
- Anderson, J.R. (1976). **Language, memory and thought**. Hillsdale: Erlbaum.
- Anderson, J.R. (1978). Arguments concerning representations for mental imagery. **Psychological Review**, 85, 249-277.
- Anderson, J.R. (1979). Further arguments concerning representations for mental imagery: Response to Hayes-Roth and Pylyshyn. **Psychological Review**, 86, 395-406.
- Anderson, J.R. (1980). **Cognitive Psychology and its implications**. San Francisco: Freeman.
- Anderson, J.R. (1983). A spreading activation theory of memory. **Journal of Verbal Learning and Verbal Behavior**, 22, 261-295.
- Anderson, J.R. & Bower, D.A. (1973). **Human associative memory**. Washington: Winston.
- Anderson, J.R. y Kline, P. (1977). Design of a production

**system.** Comunicación presentada en la Workshop on pattern-directed inference systems. Honolulu.

Anzai, Y. & Simon, H.A. (1979). The theory of learning by doing. *Psychological Review*, 86, 124-140.

Arnau, J., Honrubia, M.L. y Viader, M. (1985). **Neurociencia cognitiva: Alternativa paradigmática o metodológica?**. Comunicación presentada al III Congreso de Teoría y Metodología de las Ciencias, Gijón.

Ashby, W.R. (1954). **Design for a brain.** New York: Wiley.

- Backus, J. (1978). Can programming be liberated from the von Neumann style?. A functional style and its algebra of programs. **Communications of the ACM**, 21, 613-641.
- Bartlett, F.C. (1932). **Remembering**. Cambridge: Cambridge University Press.
- Bennett, C.H. & Landauer, R. (1985). Los límites físicos fundamentales de la computación. **Investigación y Ciencia**, 108, 30-41.
- Berg, T. (1983). **Monitoring via feedback in language production: Evidence from cut-offs**. Manuscrito no publicado.
- Bertelson, P. (1967). The time course of preparation. **Quarterly Journal of Experimental Psychology**, 19, 272-279.
- Blumenthal, R., Changeux, J.P. y Lefever, R. (1970). **Journal Membrane of Biology**, 2, 351.
- Bobrow, D. y Collins, A. (1975). **Representation and Understanding**. Studies in cognitive science. New York: Academic Press.
- Rock, J.K. (1982). Towards a cognitive psychology of syntax: Information processing contributions to sentence formulation. **Psychological Review**, 89, 1-47.
- Bower, G.H. (1975). Cognitive psychology: an introduction. En W.K. Estes (ed.): **Handbook of learning and cognitive processes**. New York: Wiley.
- Boyarski, L.L. (1967). **Current models in Biology**, 1, 39.
- Broadbent, D. (1985). A question of levels: Comment on McClelland and Rumelhart. **Journal of Experimental Psychology: General**, 114, 189-192.
- Brooks, L.R. (1984). The limits of analytic symbols: A comment on Kolers and Smythe. **Journal of Verbal Learning and Verbal Behavior**, 23, 325-327.
- Brown, J.F. (1936). On the use of mathematics in psychological theory. **Psychometrika**, 1, 77-90.



- Campbell, J. (1983). **Grammatical man**. New York: Simon & Schuster.
- Carello, C.; Turvey, M.T.; Kugler, P.N. y Shaw R.E. (1982). Inadequacies of the computer metaphor, en M.S. Gazzaniga: **Handbook of cognitive neuroscience**. New York: Plenum.
- Chase, W.G. (1973). **Visual information processing**. New York: Academic Press.
- Chase, W.G. y Clark, H.H. (1972). Mental operations in the comparison of sentences and pictures, en L. Gregg: **Cognition in learning and memory**. New York: Wiley.
- Cheng, P.W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. **Psychological Review**, 92, 414-423
- Ching Tung Lung y Dominowski, R. (1985). Effects of strategy instructions and practice on nine-dot problem solving. **Journal of Experimental Psychology (Learning, memory and cognition)**, 11, 804-811
- Chomsky, N. (1965). **Aspects of the theory of syntax**. Cambridge: MIT Press.
- Chomsky, N. (1980). Rules and representations. **The Behavioral and Brain Sciences**, 3, 1-15.
- Church, A. (1940). On the concept of random sequence. **Bulletin of the American Mathematical Society**, 46, 130-135
- Clark, H.H. (1969a). Linguistic processes in deductive reasoning. **Psychological Review**, 76, 378-404.
- Clark, H.H. (1969b). The influence of language in solving three-term series problems. **Journal of Experimental Psychology**, 82, 205-215.
- Clark, H.H. (1974). Semantics and comprehension, en T.A. Sebeok: **Current trends in linguistic**, vol 12. La Haya: Mouton.
- Cobb, L. (1978). Stochastic catastrophe models and multimodal distributions. **Behavioral Science**, 23, 360-374
- Cobb, L. (1980). Estimation theory for the cusp catastrophe theory model. **Proceedings of the Section of Survey Research Methods**. Washington, D.C.; American Statistical Society

- Cofer, C.N. (1979). **Estructura de la memoria humana.**  
Barcelona: Omega.
- Collins, A.M. y Loftus, E.F. (1975). A spreading-activation theory of semantic processing. **Psychological Review**, 82, 407-428.
- Collins, A.M. y Quillian, M.R. (1969). Retrieval time from semantic memory. **Journal of Verbal Learning and Verbal Behavior**, 8, 240-247.
- Cooper, L.A. (1973). Internal representation and transformation of random shapes. Tesis Doctoral.
- Cooper, L.A. (1975). Mental rotation of random two-dimensional shapes. **Cognitive Psychology**, 7, 20-43
- Cooper, L.A. y Shepard, R.N. (1975). Mental transformations in the discrimination of left and right hands. **Perception and Performance**, 12

- De Gandt, F. (1982). *Mathematiques et realité scientifique dans le XVII siècle, en Penser la matematicque*. Paris: Editions du Seuil.
- Debrunner, C. (1983). A two-dimensional activation cell. Working paper num. 41, Advanced Automation Research Group. Urbana, Illinois.
- Deleuze, G. (1973). *Nietzsche et la philosophie*. Paris: PUF.
- Dell, G.S. (1980). *Phonological and lexical encoding in speech production*. Tesis Doctoral no publicada.
- Dell, G.S. (1984). The representation of serial order in speech. *Journal of Experimental Psychology: Learning, memory and cognition*, 10, 222-233.
- Dell, G.S. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science*, 9, 3-24.
- Dell, G.S. y Reich, P.A. (1981). Stages in sentence production. *Journal of Verbal Learning and Verbal Behavior*, 20, 611-629.
- Dennett, D.C. (1978). *Brainstorms: Philosophical essays on mind and psychology*. Brighton: Harvester.
- Dominowski, R.L. (1981). Comment on "An examination of the alleged role of fixation in the solution of several insight problems" by Weisberg y Alba. *Journal of Experimental Psychology (General)*, 110, 199-203.
- Dretske, F.I. (1981). *Knowledge and the flow of information*. Oxford: Basil Blackwell.
- Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58, 270.

- Eich, J.M. (1982). A composite holographic associative retrieval model. *Psychological Review*, 89, 627-661.
- Ellen, P. (1982). Direction, past experience and hints in creative problem solving: Reply to Weisberg y Alba. *Journal of Experimental Psychology (General)*, 111, 316-325.
- Ericsson, K.A. y Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- Ericsson, K.A. y Simon, H.A. (1981). Sources of evidence on cognition: A historical overview, en T.V. Merluzzi et al.; *Cognitive assessment*. New York; Guilford Press.
- Evans, J.B.T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology*, 35A, 635-644.
- Evans, J.B.T. y Wason, P.C. (1976). Rationalization in a reasoning task. *British Journal of Psychology*, 67, 479-486.
- Everitt, B.S. y Hand, B.J. (1981). *Finite mixture distributions*. New York; Chapman & Hall.

- Fahlman, S.E. (1979). **NETL: A system for representing and using real-world knowledge**. Cambridge: MIT Press.
- Fahlman, S.E. (1981). Representing implicit knowledge, en G.E. Hinton y J.A. Anderson: **Parallel Models of Associative Memory**. Hillsdale: Erlbaum.
- Farah, M.J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. **Journal of Experimental Psychology (General)**, 114, 91-103.
- Feldman, J.A. (1981). A connectionist model of visual memory, en G.E. Hinton y J.A. Anderson: **Parallel Models of Associative Memory**. Hillsdale: Erlbaum.
- Feldman, J.A. y Ballard, D.H. (1982). Connectionist models and their properties. **Cognitive Science**, 6, 205-254.
- Fodor, J.A. (1965). Could meaning be an rm?. **Journal of Verbal Learning and Verbal Behavior**, 4, 73-81.
- Fodor, J.A. (1968). The appeal to tacit knowledge in psychological explanation. **Journal of Philosophy**, 65, 627-640.
- Fodor, J.A. (1975). **The language of thought**. New York: Crowell.
- Fodor, J.A. (1980). Methodological solipsism considered as a research strategy for cognitive psychology. **Behavioral and Brain Sciences**, 3, 63-73.
- Fodor, J.A. (1981a). **Representations**. Brighton: Harvester.
- Fodor, J.A. (1981b). El problema mente-cuerpo. **Investigación y Ciencia**, 54, 62-75.
- Fodor, J.A. (1983). **The modularity of mind**. Cambridge: MIT Press. Hay traducción castellana: Morata, 1986.
- Fosdick, L.D. y Osterweil, L.J. (1976). Data flow analysis of software reliability. **ACM Computing Surveys**, 8, 305-330.
- Fox, J. (1978). Continuity, concealment and visual attention, en G. Underwood (ed.): **Strategies of Information Processing**. London: Academic Press.
- Frege, G. (1918). **Logische untersuchungen**. Ernest Teil: **Der Gedanke**. Traducción castellana: Escritos lógico-semánticos, Tecnos, 1974.



Freyd, J.J. y Finke, R.A. (1984). Representational momentum. *Journal of Experimental Psychology (Learning, memory and cognition)*, 10, 126-32

Froufe, M. (1985). Introspección e informes verbales en procesamiento de información. *Estudios de Psicología*, 19-20, 135-155.

- Galotti, K.M.; Baron, J. y Sabin, J.P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models?. *Journal of Experimental Psychology: General*, 115, 16-25.
- Gibson, J.J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Giorello, G. y Morini, S. (1980). *Parabole e catastrofi*. Entrevista con René Thom. Milano: Il Saggiatore. Hay traducción castellana: Barcelona, Tusquets, 1985.
- Glass, A.L. y Holyoak, K.J. (1974/1975). Alternative conceptions of semantic memory. *Cognition*, 3, 313-339.
- Goodman, N. (1968). *Languages of art: an approach to a theory of symbols*. New York: Mac-Graw-Hill.
- Goodwin, B. (1976). La división celular y la fusión de las ideas, en C.H. Waddington: *Hacia una biología teórica*. Madrid: Alianza Editorial.
- Greene, P.H. (1962a). On looking for neural networks and cell assemblies that underlie behavior. I. A mathematical model. *The Bulletin of Mathematical Biophysics*, 24, 247-275.
- Greene, P.H. (1962b). On looking for neural networks and cell assemblies that underlie behavior. II. Neural realization of the mathematical model. *The Bulletin of Mathematical Biophysics*, 24, 395-411.
- Grice, G.R. (1968). Stimulus intensity and response evocation. *Psychological Review*, 75, 359-373
- Grossberg, S. (1980). How does the brain build a cognitive code?. *Psychological Review*, 87, 1-51.

- Harthong, S. (1984). Le moiré. **Advances in applied mathematics**, 2, 24.
- Haugeland, J. (1978). The nature and the plausibility of cognitivism. **Behavioral and Brain Sciences**, 2, 215-260.
- Haugeland, J. (1981). **Mind design**. Vermont: Bradford.
- Hayes-Roth, F. (1979). Distinguishing theories of representation. **Psychological Review**, 86, 376-382.
- Hayes-Roth, F. y Lesser, V.R. (1977). Focus of attention in the Hearsay-III speech understanding system. **Fifth International Joint Conference on Artificial Intelligence**. Cambridge, Mass.
- Hebb, D.O. (1949). **The organization of behavior**. New York: Wiley.
- Hebb, D.O. (1980). The structure of thought. En P.W. Juszczyk y R.M. Klein: **The nature of thought**. Hillsdale: Erlbaum
- Heil, J. (1980). Computation, cognition and representation. **The behavioral and brain sciences**, 3, 139
- Hendrix, G.G. (1979). Encoding knowledge in partitioned networks. En N.V. Findler (ed.): **Associative networks representation and use of knowledge by computers**. New York: Academic Press.
- Hermann, D.J.; Chaffin, R. y Daniel, M. (1984). **Three attributes of antonymy**. Manuscrito no publicado
- Hillis, W.D. (1981). The connection machine (computer architecture for the new wave). **AI MEMO 646**. MIT AI Lab. Cambridge, Mass.
- Hinton, G.E. (1979). Some demonstrations of the effects of structural descriptions in mental imagery. **Cognitive Science**, 3, 231-250.
- Hinton, G.e. (1981). Implementing semantic networks in parallel hardware. En G.E. Hinton y J.A. Anderson: **Parallel models of associative memory**. Hillsdale: Erlbaum.
- Hochberg, J.E. (1964). **Perception**. Englewood Cliffs: Prentice-Hall.
- Hockley, W.E. (1982). Retrieval processes in continuous recognition. **Journal of Experimental Psychology: Learning, memory and cognition**, 8, 497-512.

Hockley, W.E. (1984). Analysis of response time distributions in the study of cognitive processes. **Journal of Experimental Psychology: Learning, memory and cognition**, 10, 598-615.

Hockley, W.E. y Corballis, M.C. (1982). Test of serial scanning in item recognition. **Canadian Journal of Psychology**, 36, 189-212.

Hopcroft, J.E. (1984). Máquinas de Turing. **Investigación y Ciencia**, 94, 8-19.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences**

Huckfeldt, R.R.; Kohfeld, C.W. y Likens, T.W. (1982). **Dynamic modeling. An introduction**. Beverly Hills: Sage.

- Inhelder, B. y Piaget, J. (1979). Procedures et structures. **Archives de Psychologie**, 181, 165-176.
- Isnard, C.A. y Zeeman, E.C. (1976). Some models from catastrophe theory in the social sciences. En L. Collins (Ed.): **The use of models in the social sciences**. London: Tavistock.
- Iwasaki, Y. y Simon, H.A. (1986a). Causality in device behavior. **Artificial Intelligence**, 29, 3-32.
- Iwasaki, Y. y Simon, H.A. (1986b). Theories of causal ordering: Reply to de Kleer and Brown. **Artificial Intelligence**, 29, 63-72.



- Jacoby, L.L. (1983a). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, memory and cognition*, 9, 21-38.
- Jacoby, L.L. (1983b). Remembering the data: Analyzing interaction processes in reading. *Journal of Verbal Learning and Verbal Behavior*, 22, 485-508.
- Jacquette, D. (1985). Logical behaviorism and the simulation of mental episodes. *The Journal of Mind and Behavior*, 6, 325-332.
- Johnson, P. The functional equivalence of imagery and movement. *Quarterly Journal of Experimental Psychology*, 34A, 349-365
- Johnson-Laird, P.N. (1983). *Mental models*. Cambridge: Cambridge University Press.
- Johnson-Laird, P.N.; Hermann, D.J. y Chaffin, R. (1984). Only connections: A critique of semantic networks. *Psychological Bulletin*, 96, 292-315.
- Johnson-Laird, P.N. y Wason, P.C. (1977). *Thinking. Readings in cognitive science*. Cambridge: Cambridge University Press.
- Jusczyk, P.W. y Klein, R.M. (1980). *The nature of thought*. Hillsdale: Erlbaum

- Kintsch, W. (1974). **The representation of meaning in memory**. Hillsdale: Erlbaum.
- Kintsch, W. (1976). Memory for prose. En D.N. Cofer: **The structure of human memory**. San Francisco: Freeman. Hay traducción castellana: Barcelona, Omega, 1979.
- Kirkpatrick, S.; Gelatt, C.D. y Vecchi, M.P. (1983). Optimization by simulated annealing. **Science**, 220, 671-680.
- Klahr, D. (1980). Information processing models of intellectual development. En R. Kluwe y H. Spada: **Developmental Models of Thinking**. London: Academic Press.
- Klahr, D. (1983). Non monotone assessment of monotone development. En S. Strauss y R. Stay (eds.): **U-shaped behavioral growth**. New York: Academic Press.
- Klahr, D. y Wallace, J.G. (1976): **Cognitive development: An information-processing view**. Hillsdale: Erlbaum.
- Kleer, J. de y Brown, J.S. (1986). Theories of causal ordering. **Artificial Intelligence**, 29, 33-61.
- Kleinrock, L. (1975). **Quening systems**. New York: Wiley.
- Knapp, A.G. y Anderson, J.A. (1984). Theory of categorization based on distributed memory storage. **Journal of Experimental Psychology (Learning, memory and cognition)**, 10, 616-637
- Köhler, W. (1947). **Gestalt Psychology**. New York: Liveright.
- Köhler, W. (1955). **Dinámica en Psicología**. Buenos Aires: Paidós.
- Kohonen, T. (1977). **Associative memory. A system-theoretical approach**. Berlin: Springer-Verlag.
- Kolers, P.A. y Ducknicky, R.L. (1985). Discontinuity in cognitive skill. **Journal of Experimental Psychology: Learning, memory and cognition**, 11, 655-674.
- Kolers, P.A. y Smythe, W.E. (1979). Images, symbols and skills. **Canadian Journal of Psychology**, 33, 158-184.
- Kolers, P.A. y Smythe, W.E. (1984). Symbol manipulation: Alternatives to the computational view of mind. **Journal of Verbal Learning and Verbal Behavior**, 23, 289-314.

Kosslyn, S. (1980). *Image and mind*. Cambridge, Mass.:  
Harvard University Press.

Kuffer, S.W. y Nichols, J.G. (1976). *From neurons to brain*.  
Sunderland: Sinauer Associates.

- Lachman, J. y Lachman, R. (1979). Theories of memory organization and human evolution. En C.R. Puff (Ed.): **Memory organization and structure**. New York: Academic Press.
- Lashley, K.S. (1942). The problem of cerebral organization in vision. En H. Kluver: **Biological Symposia, VII**. Lancaster: Jaques Cattell Press.
- Lashley, K.S. (1950). In search of the engram. En **Symposia of the Society for Experimental Biology, 4**. New York: Academic Press.
- Levine, M. (1975). **A cognitive theory of learning**. Hillsdale: Erlbaum
- Lewin, K. (1934). Environmental forces. En C. Murchison: **Handbook of child psychology**. Worcester: Clark University Press.
- Lewin, K. (1935). **A dynamic theory of personality**. New York: McGraw-Hill.
- Lewin, K. (1936). **Principles of topological psychology**. New York: McGraw-Hill.
- Lewin, K. (1938). The conceptual representation and measurement of psychological forces. **University of Iowa Contr. Psychol. Theory, 1**, 1-247
- Lewin, K. (1951). **Field theory in social science**. New York: Harper.
- Lewis, C.H. (1978). **Production system models of practices effects**. Documento Universidad de Michigan.
- Lurié, D. y Wagensberg, J. (1983). On biomass diversity in ecology. **Bulletin of Mathematical Biology, 45**.
- Lutz, R. y Goze, M. (1981). **Lecture notes in mathematics**. Berlin: Springer-Verlag.

- MacKay, D.G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8, 323-350.
- Malapeira, J.M. (1987). **Análisis teórico-metodológico de la psicofísica, la cronoscopia y la teoría de la detección de señales.** Tesis Doctoral.
- Maloney, J.C. (1985). Methodological solipsism reconsidered as a research strategy in cognitive psychology. *Philosophy of Science*, 52, 451-469.
- Mandler, G. (1967). Verbal learning. En G. Mandler et al.: **New directions in psychology, III.** New York: Holt, Rinehart y Winston.
- Marble, H.C. (1960). **The hand. A manual and atlas for the general surgeon.** Philadelphia: Saunders.
- Margalef, R. (1968). **Perspectives in ecological theory.** Chicago; University of Chicago Press. Hay traducción castellana: Blume.
- Margalef, R. (1972). **Homage to Evelyn Hutchinson, or why there is an unpper limit to diversity.** University of Connecticut.
- Marr, D. (1977). Artificial intelligence. A personal view. *Artificial Intelligence*, 9, 37-48.
- Marr, D. (1982). **Vision.** San Francisco: Freeman.
- McAlister, L.L. (1976). **The philosophy of Brentano.** London: Duckworth.
- McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287-330.
- McClelland, J.L. y Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception. Part I. An account of basic findings. *Psychological Review*, 88, 375-407.
- McClelland, J.L. y Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159-188.
- McCulloch, W.S. y Pitts, W.H. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.



- McDermott, J. y Forgy, C. (1978). Production system conflict resolution strategies. En D.A. Waterman y f. Hayes-Roth, (Eds.): **Pattern-directed inference systems**. New York: Academic Press.
- McDowd, J.M. y Murdock, B.B. (1986). Mathematical models of memory and the problem of stimulus variation: A comparison of MINERVA 2 and TODAM. **Acta Psychologica**, 62, 177-188.
- McGill, J.W. (1963). Stochastic latency mechanisms. En R.D. Luce, R.R. Bush y E. Galanter: **Handbook of mathematical psychology**. New York: Wiley.
- Metcalfe, J.A. y Murdock, B.B. (1981). An encoding and retrieval of single trial free recall. **Journal of Verbal Learning and Verbal Behavior**, 20, 161-189.
- Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A. y Teller, E. (1953). Equation of state calculations for fast computing machines. **Journal of Chemical Physics**, 6, 1087.
- Meyer, D.E. y Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words. **Journal of Experimental Psychology**, 90, 227-234.
- Meyer, D.E.; Yantis, S.; Osman, A. y Smith, J.E.K. (1984). Discrete versus continuous models of response preparation: A reaction-time analysis. En S. Kornblum y J. Requin: **Preparatory states and processes**. Hillsdale: Erlbaum.
- Meyer, D.E.; Yantis, S.; Osman, A. y Smith, J.E.K. (1985). Temporal properties of human information processing: Tests of Discrete versus Continuous models. **Cognitive Psychology**, 17, 445-518.
- Minsky, M. (1975). A framework for representing knowledge. En P. H. Winston: **The psychology of computer vision**. New York: McGraw-Hill.
- Minsky, M. (1979). The society theory of thinking. En P.H. Winston y R.H. Brown: **Artificial Intelligence: An MIT Perspective**. Cambridge: MIT Press.
- Minsky, M. y Papert, S. (1969). **Perceptrons**. Cambridge: MIT Press.
- Molenaar, P. (1986). Issues with a rule sampling theory of conservation learning from a structuralist point of view. **Human Development**, 29, 137-144.

- Molenaar, P. y Oppenheimer, L. (1985). Dynamic models of development and the mechanistic-organismic controversy. **New Ideas in Psychology**, 3, 233-242.
- Monod, J. (1981). **El azar y la necesidad**. Barcelona: Tusquets.
- Morris, P.E. y Hampson, P.J. (1983). **Imagery and consciousness**. London: Academic Press.
- Morrison, P. y Morrison, E. (1961). **Phillip Babbage and his calculating engines**. New York: Dover Publications.
- Morton, J. (1969). The interaction of information in word recognition. **Psychological Review**, 76, 165-178.
- Morton, J. (1979). Facilitation of word recognition: Experiments causing change in the logogen model. En P. A. Kolers, M.E. Wrolstad y H. Bowma: **Processing of visible lenguaje**. New York: Plenum.
- Morton, J. (1981). The status of information processing models of language. **Philosophical transactions of the Royal Society of London**, 295, 387-395.
- Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. **Psychological Review**, 89, 609-626.
- Murdock, B.B. (1983). A distributed memory model for serial-order information. **Psychological Review**, 90, 316-338.

- Neely, J.H. (1976). Semantic priming and retrieval from lexical memory. **Memory & Cognition**, 4, 648-654.
- Neves, D.M. y Anderson, J.R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. En J.R. Anderson: **Cognitive skills and their acquisition**. Hillsdale: Erlbaum.
- Newell, A. (1973). Production systems. Models of control structures. En W. Chase: **Visual information processing**. New York: Academic Press.
- Newell, A. y Simon, H.A. (1972). **Human problem solving**. Englewood Cliffs: Prentice-Hall.
- Nisbett, R.E. y Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. **Psychological Review**, 84, 231-259.
- Norman, D.A. y rumelhart, D.E. (1975). **Explorations in cognition**. San Francisco: Freeman.

- Pachella, R.G.; Smith, J.E.K. y Stanovich, K.E. (1978). Qualitative error analysis and speeded classification. En N.J. Castellan y F. Restle: **Cognitive theory**, vol. 3fn. Hillsdale; Erlbaum.
- Paivio, A. (1971). **Imagery and verbal processes**. New York: Holt, Rinehart & Winston.
- Paivio, A. (1977). Images, propositions and knowledge. En J.M. Nicholas: **Images, perception and knowledge**. Dordrecht: Reidel.
- Paivio, A. (1978). Mental comparisons involving abstract attributes. **Memory and Cognition**, 6, 199-208
- Paivio, A. (1979). The relationship between verbal and perceptual codes. En E.C. Carterette y M.P. Friedman: **Handbook of perception**. New York: Academic Press.
- Paivio, A. (1983). The empirical case for dual coding. En: J.C. Yuille: **Imagery, memory and cognition**. Hillsdale: Erlbaum
- Paivio, A. y Begg, I. (1981). **Psychology of language**. Englewood Cliffs: Prentice-Hall.
- Paivio, A. y Csapo, K. (1969). Concrete-image and verbal memory codes. **Journal of Experimental Psychology**, 80, 279-285
- Paivio, A. y Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding?. **Cognitive Psychology**, 5, 176-206
- Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. **Psychological Review**, 91, 281-294
- Poincaré, H. (1914). **Le hasard**. Paris: Flammarion
- Pollack, J. (1982). An activation/inhibition network WLSI cell. Working paper 31. **Coordinated Science Laboratory**, Urbana, Illinois
- Poston, T. y Stewart, I.N. (1978). **Catastrophe theory and its applications**. London: Pitman
- Pribram, K.H. (1975). Toward a holonomic theory of perception. En: **Gestalttheorie in der moderne Psychologie**
- Pribram, K.H. (1977). **Languages of the brain: Experimental paradoxes and principles in neuropsychology**. Englewood Cliffs: Prentice-Hall

- Pribram, K.H. y Martín Ramirez, J. (1980). **Cerebro, mente y holograma**. Madrid: Alhambra
- Prigogine, I. (1983). **¿Tan sólo una ilusión?**. Barcelona: Tusquets
- Prigogine, I. y Stengers, I. (1979). **La nouvelle alliance**. Paris: Gallimard
- Pylyshyn, Z.W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. **Psychological Bulletin**, 80, 1-24.
- Pylyshyn, Z.W. (1977). What does it take to bootstrap a language?. En J. Macnamara: **Language learning and thought**. New York: Academic Press.
- Pylyshyn, Z.W. (1978). Computational models and empirical constraints. **Behavioral and Brain Sciences**, 1, 93-127.
- Pylyshyn, Z.W. (1979). Validating computational models. A critique of Anderson's indeterminacy of representation claim. **Psychological Review**, 86, 384-394.
- Pylyshyn, Z.W. (1980). Cognitive representation and the process-architecture distinction. **The Behavioral and Brain Sciences**, 3, 154-169.
- Pylyshyn, Z.W. (1981). The imagery debate: Analogue media versus tacit knowledge. **Psychological Review**, 88, 16-45.
- Pylyshyn, Z.W. (1983). La naturaleza simbólica de las representaciones mentales. En M.V. Sebastián: **Lecturas de psicología de la memoria**. Madrid: Alianza.
- Pylyshyn, Z.W. (1986). **Computation and cognition**. Cambridge: MIT Press.



- Rashevsky, N. (1948). **Mathematical Biophysics**. Chicago: University of Chicago Press.
- Ratliff, F. (1983). Comunicación presentada en la New York Association for the Blind.
- Ratcliff, R. (1978). A theory of memory retrieval. **Psychological Review**, 85, 59-108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. **Psychological Bulletin**, 86, 446-461.
- Ratcliff, R. y Murdock, B.B. (1976). Retrieval processes in recognition memory. **Psychological Review**, 83, 190-214.
- Reed, S.K. (1984). Estimating answers to algebra word problems. **Journal of Experimental Psychology (Learning, Memory and Cognition)**, 10, 778-790.
- Reed, S.K.; Dempster, A. y Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. **Journal of Experimental Psychology (Learning, Memory and Cognition)**, 11, 106-125.
- Rivière, A. (1984). Sobre la multiplicidad de las representaciones. Un viaje por los vericuetos de los lenguajes del pensamiento. Symposium sobre Actividad Humana y Procesos Cognitivos. Madrid.
- Rivière, A. (1986). **Razonamiento y representación**. Madrid: Siglo XXI.
- Robinson, A. (1974). **Non standard analysis**. North Holland.
- Rosen, R. (1986). Causal structures in brains and machines. **International Journal of General Systems**, 12, 107-126.
- Rosenberg, S. y Simon, H.A. (1977). Modeling semantic memory: effects of presenting semantic information in different modalities. **Cognitive Psychology**, 9, 293-325.
- Rosenblatt, F. (1961). **Principles of neurodynamics: Perceptrons and the theory of brain mechanisms**. Washington, D.C.: Spartan.
- Roth, I. y Frisby, J.P. (1986). **Perception and representation. A cognitive approach**. Philadelphia: Open University Press.
- Rowe, H.A.H. (1985). **Problem solving and intelligence**. Hillsdale: Erlbaum.

- Rumelhart, D.E. y McClelland, J.L. (1985). Levels indeed!. A response to Broadbent. **Journal of Experimental Psychology: General**, 114, 193-197.
- Rumelhart, D.E. y Norman, D.A. (1981). Introducción a la obra de Hinton y Anderson: **Parallel models of associative memory**. Hillsdale: Erlbaum.
- Russell, B. (1903). **The principles of mathematics**. Cambridge
- Rychener, M. y Newell, A. (1978). An intractable production system: initial design issues. En D.A. Waterman y F. Hayes-Roth: **Pattern-directed inference systems**. New York: Academic Press

- Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 2, 437-442.
- Schank, R.C. (1973). Identification of conceptualization underlying natural language. En R.C. Schank y K.M. Colby: *Computer models of thought and language*. San Francisco: Freeman.
- Schank, R.C. (1975). *Conceptual information processing*. Amsterdam: North Holland.
- Schank, R.C. (1979). El papel de la memoria en el procesamiento del lenguaje. En C.N. Cofer: *Estructura de la memoria humana*. Barcelona: Omega.
- Schank, R.C. (1981). Language and memory. En D.A. Norman: *Perspectives on cognitive science*. Hillsdale: Erlbaum.
- Schank, R.C. (1982). *Dynamic memory*. Cambridge: Cambridge University Press.
- Schank, R.C. y Abelson, R.P. (1977). *Scripts, plans, goal and understanding*. Hillsdale: Erlbaum.
- Scheerer, M. (1963). Problem solving. *Scientific American*, Abril 1963.
- Schrödinger, E. (1945). *¿Qué es la vida?*; traducción castellana; Tusquets, 1983
- Sekuler, R., Rubin, E. y Armstrong, R. (1971). Processing numerical information: A choice time analysis. *Journal of Experimental Psychology*, 90, 75-80
- Shepard, R.N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 27, 125-140.
- Shepard, R.N. (1975). Form, formation and transformation of representations. En R.L. Solso: *Information processing and cognition*. Hillsdale: Erlbaum
- Shepard, R.N. (1978). The mental image. *American Psychologist*, 33, 125-137
- Shepard, R.N. (1984). Ecological constraints on internal representation. Resonant kinematics of perceiving, imagining, thinking and dreaming. *Psychological Review*, 91, 417-447.
- Shepard, R.N. y Chapman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1, 1-18

- Shepard, R.N. y Feng, C. (1972). A chronometric study of mental paper folding. **Cognitive Psychology**, 3, 228-243
- Shepard, R.N., Kilpatrick, D.W. y Cunningham, J.F. (1975). The internal representation of numbers. **Cognitive Psychology**, 5
- Shiffrin, R.M. y Schneider, W. (1977). Controlled and automatic human information processing. **Psychological Review**, 84, 127-190.
- Simmons, G.F. (1967). **Topology and modern analysis**. New York: McGraw-Hill
- Simon, H.A. (1975). The functional equivalence of problem solving skills. **Cognitive Psychology**, 7, 268-288.
- Simon, H.A. (1978). Information processing theory of human problem solving. En W.K. Estes (Ed.): **Handbook of learning and cognitive processes**. Vol 5. Hillsdale: Erlbaum
- Sternberg, R.J. (1979). The nature of mental abilities. **American Psychologist**, 34, 214-230.
- Sternberg, S. (1966). High speed scanning in human memory. **Science**, 153, 652-654.
- Sternberg, S. (1969). The discovery of processing stages: extensions of Donders method. **Acta Psychologica**, 30, 276-315.
- Sternberg, S. (1975). Memory scanning: New findings and current controversies. **Quarterly Journal of Experimental Psychology**, 27, 1-32.
- Stewart, I.N. y Peregoy, P.L. (1983). Catastrophe theory modeling in psychology. **Psychological Bulletin**, 94, 336-362.
- Stich, S. (1984). **Folk psychology and cognitive science. The case against belief**. Cambridge: MIT Press.
- Stinessen, L. (1985). The influence of verbalization on problem solving. **Scandinavian Journal of Psychology**, 26, 342-347.
- Sussman, H.J. y Zahler, R.S. (1978). Catastrophe theory as applied to the social and biological sciences: a critique. **Synthese**, 37, 117-216.
- Sutton, R.S. y Barto, A.G. (1981). Toward a modern theory of adaptative networks: Expectation and prediction. **Psychological Review**, 88, 135-170.

Sweller, J., Mawer, R.F. y Ward, M.R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology (General)*, 112, 639-661



- Thom, R. (1972). *Stabilité structurelle et morphogenèse*. Reading: Benjamin.
- Thom, R. (1980). *Modèles matemáticos de la morphogenèse*. Paris: Bourgois.
- Thom, R. (1982). Matemáticas y teorización científica. En: *Pensar la Matemática*. Barcelona: Tusquets.
- Thom, R. (1985). *Parábolas y catástrofes*. Barcelona: Tusquets.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge: MIT Press.
- Vega, M. de (1981). Una exploración de los metapostulados de la psicología contemporánea: el logicismo. *Análisis y Modificación de Conducta*, 7, 345-375.
- Vega, M. de (1984). *Introducción a la psicología cognitiva*. Madrid: Alianza.
- Waddington, C.H. (1976). *Hacia una biología teórica*. Madrid: Alianza.
- Wagensberg, J. (1985). *Ideas sobre la complejidad del mundo*. Barcelona: Tusquets.
- Waltz, D.L. y Pollack, J.B. (1985). Massively parallel parsing. A strongly interactive model of natural language interpretation. *Cognitive Science*, 9, 51-74.
- Wason, P.C. y Evans, J.S.B. (1975). Dual processes in reasoning?. *Cognition*, 3, 141-154.
- Weisberg, R.W. y Alba, J.W. (1981a). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology (General)*, 110 162-192.
- Weisberg, R.W. y Alba, J.W. (1981b). Gestalt theory, insight and past experience: Reply to Dominowski. *Journal of Experimental Psychology*, 110, 193-198
- Weisberg, R.W. y Alba, J.W. (1982). Problem solving is not like perception: More on Gestalt theory. *Journal of Experimental Psychology (General)*, 111, 326-330.
- Wickelgren, W.A. (1972). Coding, retrieval and dynamics of multirace associative memory. En L.W. Gregg: *Cognition in learning and memory*. New York: Wiley.

- Wickens, T.D. (1982). **Models for behavior**. San Francisco: Freeman.
- Wilson, K.V. (1980). **From associations to structure**. Amsterdam: North Holland.
- Wilson, H.R. y Cowan, J.D. (1971). **Excitatory and inhibitory interactions in localized populations of model neurons**. Chicago: Chicago University Press
- Windelband, W. (1894). **geschichte und Naturwissenschaft**. Tübingen
- Winograd, T. (1971). **Procedures as a representation for data in a computer program for understanding natural language**. Cambridge: MIT Press.
- Winograd, T. (1984). **Comprensión del lenguaje**. *Investigación y Ciencia*, 57, 67-88.
- Wolman, B.B (1960). **Teorías y sistemas contemporáneos en Psicología**. (Traducción castellana: Martínez Roca, 1968).
- Yuille, J.C. (1983). **Imagery, memory and cognition**. Hillsdale: Erlbaum.
- Zeeman, E.C. (1977). **Catastrophe theory: Selected papers 1972-1977**. Reading: Addison-Wesley.