

Sistemas expertos y psicología cognitiva: una visión general

PABLO ADARRAGA y JOSÉ L. ZACCAGNINI

Universidad Autónoma de Madrid



Resumen

El presente trabajo intenta ofrecer una visión sintética del tema de los sistemas expertos que trata de hacer hincapié en aquellos aspectos de este campo que pudieran resultar más relevantes para el psicólogo interesado en la cognición. Se traza un breve esbozo del marco científico-técnico de la inteligencia artificial en general, a partir del cual se intenta caracterizar a los sistemas expertos desde dos puntos de vista principales: en primer lugar, el tipo de problemas y tareas a los que se destinan dichos sistemas y, en segundo lugar, la estructura computacional subyacente que ha permitido abordarlos con éxito. Se discute la relación histórico-científica entre la inteligencia artificial y la psicología cognitiva, observándose que se ha producido un flujo constante y bidireccional de conceptos teóricos, métodos y problemas conceptuales de una a otra, especialmente en el actual marco interdisciplinar de la ciencia cognitiva. Dado este contexto, se argumenta que es muy posible que los sistemas expertos aporten novedades conceptuales y metodológicas significativas para la Psicología. Por último, se especula sobre algunas cuestiones en las que podría tener lugar tal contribución.

Abstract

The present paper introduces a brief survey of the field of expert systems, biased toward the topics that appear as more relevant to psychologists concerned with cognition. A sketch is presented of the scientific and technical framework of general artificial intelligence, and, using it as a background, an attempt is made to characterize expert systems from two main standpoints: first, the sort of problems and tasks that such systems are assigned and, second, the underlying computational structures that enable them to deal with those problems and tasks successfully. The historical and scientific relationship between artificial intelligence and cognitive psychology is discussed, nothing that there has been a regular interchange of theoretical concepts, methods and problems, specially within the interdisciplinary framework of cognitive science. It is subsequently claimed that expert systems might well entail significant contributions to psychology. Finally, some speculations are made about the issues in which such contributions might take place.

Dirección del autor: Pablo Adarraga. Departamento de Psicología Básica, Social y Metodología. Facultad de Psicología. Universidad Autónoma de Madrid. Ciudad Universitaria de Canto Blanco. Telf. 39 75 00. 28049 Madrid.

INTRODUCCION

Las fronteras entre disciplinas científicas que tratan temas afines son siempre fascinantes, pero también a menudo confusas. El caso de la psicología cognitiva (PC) y la inteligencia artificial (IA) no es ninguna excepción: las dos tienen como objeto de estudio común el comportamiento inteligente, y existe una compleja zona de solapamiento o intersección entre ambas, con todos los problemas conceptuales que ello supone. Tratando de poner un poco de orden en este sentido, se ha utilizado a menudo la idea de que la diferencia entre la IA y ciertos desarrollos de la PC es una cuestión que se plantea en el nivel de los objetivos básicos. En consecuencia, se ha manejado una distinción entre IA y simulación psicológica. *La IA propiamente dicha*, según tal distinción, perseguiría el objetivo de diseñar sistemas computacionales con *prestaciones* equiparables a las del intelecto humano, independientemente de que éstas se logren por procesos parecidos a los humanos o no. Por tanto, la IA sería esencialmente ajena a la Psicología: por más que ambas traten problemas coincidentes, el enfoque es absolutamente diferente («cómo consiguen los sujetos hacer X» frente a «cómo puedo conseguir que mi computador haga X», siendo X cualquier tarea con demandas intelectuales). Por otra parte, se habla de *simulación*: una técnica metodológica consistente en construir programas de computador que reproduzcan los componentes postulados por una teoría o modelo *psicológico* de una función cognitiva. La simulación supondría, pues, una manera de expresar una teoría psicológica que resulta muy exigente en términos de precisión y coherencia, y que permite formas peculiares de contrastación de la misma (cfr. De Vega, 1982). En definitiva, se trataría de una metodología de trabajo utilizable por la PC, subordinada en cuanto tal a lo que el investigador quiera hacer con ella, y potencialmente alternativa o complementaria a las tradicionales técnicas experimentales.

Però en los últimos años la situación se ha vuelto más compleja. En el contexto del pujante desarrollo de la IA aparecen expresiones como Ingeniería del Conocimiento, Arquitectura Cognitiva, Sistemas Basados en Conocimiento, Sistemas Expertos, etc. Se trata frecuentemente de rótulos que sugieren temas relacionados con la PC, pero no suele quedar claro en la literatura hasta qué punto hacen referencia a contenidos verdaderamente significativos para la misma, ni resulta evidente si la ya de por sí frágil distinción «IA vs. Simulación» sigue siendo útil para determinar el estatus de relevancia que podemos atribuirles desde nuestra perspectiva. Por otra parte, es obvio que la aceptación de que la IA y la PC sean fundamentalmente distintas no invalida el hecho de que históricamente se hayan producido numerosas aportaciones muy significativas de cada una a la otra (Newell, 1973a; De Vega, 1982, 1984, y Delclaux, 1982).

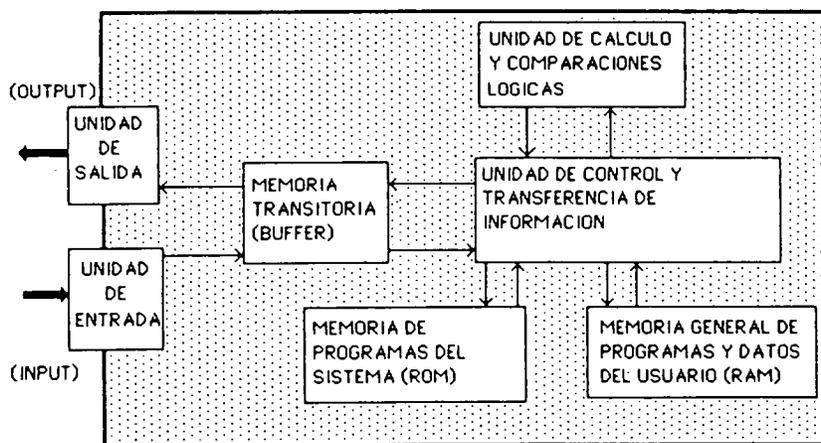
En línea con esa idea, el objetivo del presente trabajo es tratar de explorar, desde la perspectiva de la PC, uno de los nuevos desarrollos de la IA: los *sistemas expertos*. La elección no es arbitraria: seguramente se trata del campo que mayor cantidad de literatura, investigación y aplicaciones prácticas ha generado en el contexto de la IA reciente. Y, sobre todo, ha originado novedades conceptuales importantes, respecto a las cuales merece la pena preguntarse hasta qué punto suponen ideas ya convergentes, ya divergentes, potencialmente útiles, o bien completamente ajenas a la PC.

Una primera cuestión que complica un tanto la exploración de esas posibles relaciones es que la PC es una disciplina difícil de acotar en sí misma de forma precisa. Como señala Rivièrè (1987), no se trata de un «concepto» bien definido (en la línea de los estudios clásicos de formación de conceptos), sino de una «categoría natural» (Rosch, 1978; Rosh y Mervis, 1975), definida por un conjunto de atributos de desigual relevancia («validez de señal») y con elementos de variada representatividad («tipicidad»). Como consecuencia, la calificación de «cognitiva», según la acepción que se le quiera dar, puede referirse a un enfoque muy particular y relativamente bien delimitado o, por el contrario, a casi toda la Psicología actual (para una discusión de este punto, véase por ejemplo Zaccagnini y Delclaux, 1982; Zaccagnini y Morales, 1985, y Zaccagnini, 1984). Por ello, lo primero que es obligado hacer en un artículo como el presente es declarar explícitamente qué significado se le atribuye. Al hablar aquí de *Psicología Cognitiva* vamos a referirnos a cualquier corriente o sistema teórico de la Psicología científica que, de forma explícita o implícita, asuma que el comportamiento humano está mediado y controlado al menos en parte por entidades o fenómenos, de carácter representacional, internos al organismo, y sin referencia a los cuales es imposible una explicación psicológica satisfactoria de dicho comportamiento; presuponiendo también que dichas entidades o fenómenos internos poseen propiedades emergentes no reductibles a las del medio físico-orgánico en que tienen lugar, exigiendo para su estudio y comprensión un plano o nivel de análisis propio y diferenciado (Rivièrè, 1986, 1987). Es decir, la PC implicaría un doble metapostulado (en el sentido de De Vega, 1981, 1983): un presupuesto internalista o *mentalista*, y un presupuesto *emergentista*. Al atribuir este significado a «Psicología Cognitiva» creemos coincidir con la opinión de la gran mayoría de los autores (Gardner, 1985, y Rivièrè, 1987). Nótese que este criterio excluye de la PC, por ejemplo, las explicaciones conductistas o las fisicalistas y, sin embargo, admite enfoques «atípicos» o periféricos de la categoría como la Escuela Sociohistórica, la Cognición Social, el neoconexionismo o las aproximaciones ecológicas [pese al radicalismo de algunas de estas últimas (Gibson, 1979), lo cierto es que la mayoría terminan por postular entidades internas irreductibles (vg. Neisser, 1976)]. Como hemos sostenido en otros lugares (Zaccagnini, 1984 y Adarraga, 1986), es posible entender el nivel cognitivo de explicación como un nivel sistémico que, aun poseyendo sus propiedades diferenciales, esté sometido a restricciones, demandas e influencias por parte de otros niveles, como el sociocultural o el fisiológico. Así pues, es posible que el «solipsismo» o la ignorancia de lo biológico que tanto (y tan certeramente) se han criticado a la PC, no sean características necesarias, intrínsecas a ésta.

Un subconjunto especialmente representativo de la PC así entendida será la *Psicología del Procesamiento de Información* (PPI), la cual se define en virtud de un metapostulado adicional más específico, que se ha dado en llamar la analogía o metáfora del computador. Según este presupuesto, esos mediadores internos (actividad mental) son científicamente descriptibles, en el lenguaje conceptual de las ciencias del cómputo, como procesos que se efectúan sobre información mediante manipulación de símbolos (Newell y Simon, 1972; Zaccagnini y Delclaux, 1982; De Vega, 1982, 1983, etc.). En la figura 1 se representa un esquema de la estructura funcional de un siste-

ma general de procesamiento de información, que corresponde, en líneas generales, a la segmentación epistemológica de los fenómenos mentales que ha guiado la investigación bajo el enfoque de la PPI (Ruiz-Vargas y Zaccagnini, 1987). Sin embargo, no sería correcto afirmar que dicha investigación ha asumido de manera uniforme tal marco general. En la práctica, en torno a la analogía del ordenador se arremolinan subenfoques o interpretaciones de muy diversa índole: algunos autores la tratan simplemente como una metáfora heurísticamente útil, mientras otros la asumen como hipótesis general sobre la naturaleza de la actividad mental; se adoptan posiciones epistemológicas de corte realista («los sujetos computan de hecho información») o bien de tipo más «instrumental» («es útil describir los fenómenos mentales como si fueran manipulación computacional de información»); entre los intereses de los investigadores se encuentran en combinaciones diversas la explicación de los procesos mentales específicamente humanos y la resolución de las funciones cognitivas en abstracto e independientemente del procesador concreto que las realice; estas opciones se expresan abiertamente o se asumen de forma tácita, y arrojan producciones teóricas de muy diverso éxito en cuanto a su coherencia, significación empírica o relevancia para una «psicología natural» (en el sentido de Rivière, 1987). A fin de tratar de poner cierto orden en este panorama, se han utilizado distinciones como la «versión fuerte» y la «versión débil» de la analogía del ordenador (muy relacionada con la dicotomía IA-simulación a que antes nos referíamos), aunque los propios autores que se sirven de estos conceptos señalan sus limitaciones a la hora de clasificar trabajos concretos (De Vega, 1982).

FIGURA 1



Las partes funcionales de un computador como dispositivo general de procesamiento de información (adaptado de Ruiz Vargas y Zaccagnini, 1987).

En cualquier caso, para los fines que aquí nos proponemos no será necesario profundizar más en esta cuestión (véase el lúcido análisis de Rivière, *op. cit.*). Únicamente subrayar explícitamente que en este trabajo no identificamos PC y PPI, sino que asumimos una distinción todo-parte conforme a lo antedicho.

EL MARCO DE LA INTELIGENCIA ARTIFICIAL

Sería interesante explicitar, del mismo modo que hemos hecho con la PC, a qué vamos a llamar aquí Inteligencia Artificial. En la práctica, esto es más difícil, si cabe, ya que «Inteligencia Artificial» es en realidad una etiqueta que se aplica a un amplio conjunto de líneas de trabajo en informática, cuyos propósitos, bases teóricas, tradiciones de trabajo, etc., son notablemente heterogéneos. Como señala Naughton (1986), «... es más bien un estado mental —una manera particular de abordar la resolución de problemas y el uso de los computadores— que una disciplina académica rigurosamente definida» (p. 44).

A lo largo de su historia, se han dado diversas definiciones y caracterizaciones de la IA, al mismo tiempo que se ha hecho énfasis en distintos tipos de tareas a las que aplicarla (deducción automática, percepción, manejo de conocimiento...). Vamos a tratar de hacer una acotación que resulte todo lo amplia posible, procurando al mismo tiempo mantener una perspectiva sintética.

Como es bien sabido, lo que hoy llamamos Inteligencia Artificial tiene sus orígenes en los primeros años cincuenta, cuando algunos investigadores (en particular Claude Shannon) comienzan a explorar las posibilidades de los computadores en la resolución de problemas diferentes de los típicamente matemáticos (Aleksander, 1986). En un principio, las aplicaciones prácticas de los computadores se habían dirigido a utilizarlos únicamente como «supercalculadoras»: máquinas capaces de efectuar a gran velocidad complejos cálculos numéricos; pero pronto se empieza a explorar la llamativa posibilidad de dirigir sus recursos lógico-aritméticos hacia el manejo de problemas simbólicos o de razonamiento. Esta idea es, en realidad, anterior a la tecnología que en los años cuarenta permitió la construcción física de computadores; y había sido formulada explícitamente por autores como Babbage o Turing (McCorduck, 1979). Antes de terminar los cincuenta, son ya bastantes los centros que investigan la aplicación de computadores en la ejecución de diversos tipos de tareas. Principalmente, se trata de sacar partido de su capacidad para calcular las consecuencias sucesivas de secuencias de acciones o decisiones, buscando las vías más adecuadas al criterio de la tarea. Cada acción adoptada trae consigo un abanico, más o menos amplio, de continuaciones posibles, cada una de las cuales a su vez abre nuevas posibilidades. Así, la resolución de un problema se puede entender como un proceso de «exploración» en un «árbol de decisiones»: el estado inicial de la tarea se concibe como un «tronco» del cual parten diversas ramas (cada una de las posibles acciones a partir de tal estado), y así sucesivamente hasta alcanzar los estados finales, simbolizados por las ramas terminales. Planteada de esta forma, la resolución del problema o tarea supone una «exploración» del árbol, en busca de estados (ramas) terminales que cumplan determinadas condiciones. Así, si recordamos el bien conocido ejemplo del ajedrez, un determinado estado del tablero supone un «nudo» del cual parten ramas (jugadas) hacia otros estados, y el problema se reduce a buscar en cada momento el camino más conveniente según unos criterios dados.

Problemas de diferente naturaleza generan árboles de búsqueda más o menos amplios y complejos. Por ejemplo, sin salir del terreno de los jue-

gos (tema favorito de la temprana IA), en el caso del tres-en-rama el árbol es mucho menos extenso que el ajedrez. En general, sin embargo, la idea del árbol de búsqueda resulta ser una potente representación del «espacio del problema» para gran número de tipos de tareas diferentes, sean o no matemáticas. Y, lo que es fundamental, *es un formato muy adecuado para acometerlas mediante computador*: éste se ocupará de evaluar las distintas ramas accesibles en cada momento y optar por la más prometedora.

Es en 1956 cuando John McCarthy, de la Universidad de Stanford, acuña la expresión «Inteligencia Artificial» para cubrir un ya relativamente amplio conjunto de sistemas y proyectos. Muchos autores consideran —y tenemos que coincidir con ellos— que esta denominación no es idónea. Por ejemplo, Aleksander (1986) señala que «... quizá es la palabra “inteligencia” la que ha creado una cierta mística en torno a lo que de otro modo no es sino una forma obvia de resolver problemas en computadores [...] Indudablemente esta expresión [inteligencia artificial] posee unas resonancias de reemplazo del hombre que, en el contexto de un intento de introducir la IA en los lugares de trabajo, no es de gran ayuda» (p. 12). De forma más radical, Naughton (1986) considera que buena parte de la literatura sobre el tema «abusa ridículamente del término inteligencia» (p. 44). Sin embargo, la denominación coherente con lo que en esta primera época (a falta de algo mejor) sirve como criterio de «comportamiento inteligente»: un programa «posee inteligencia» si es capaz de ejecutar una función o tarea para la que normalmente se consideraría necesario un entendimiento humano; en este contexto hay que entender, por ejemplo, la famosa prueba de Turing. Este concepto de «inteligencia artificial» se ha mantenido, en líneas generales, hasta la actualidad, y aunque la caracterización de los «programas inteligentes» se ha ido perfilando con los sucesivos avances, sigue siendo esencialmente arbitraria (Haugeland, 1981). A pesar de todo, hay que señalar que en otras partes del mundo, las incursiones de los computadores en los problemas no matemáticos recibían otras denominaciones quizá menos pretenciosas, como «Informática Experimental» (Experimental Computing) en la Universidad de Edimburgo.

Independientemente de la mayor o menor fortuna de la expresión, lo cierto es que los trabajos y productos que se consideran normalmente como IA poseen rasgos comunes que permiten agruparlos e identificarlos (aunque, de nuevo, más bien como categoría natural que como concepto formal). Uno de ellos ya lo hemos señalado: se va a tratar de trabajos que lleven el uso de los computadores a terrenos cada vez más lejanos de aquellos en los que inicialmente se aplicaron; problemas para cuya resolución resultaba anteriormente indispensable el concurso de un agente humano.

A finales de los años sesenta, el desarrollo de los programas inteligentes basados en la «exploración del árbol de búsqueda» había sido ya grande. Los esfuerzos se habían centrado esencialmente en diseñar estrategias generales de resolución de problemas, e incorporarlas a programas de los que se esperaban comportamientos inteligentes en rangos arbitrariamente amplios de tipos de problemas. Ello había producido avances conceptuales importantes, pero el éxito no era completo. Los programas inteligentes de propósito general poseían un indudable interés desde el punto de vista científico-técnico y académico, pero no acababan de resultar verdaderamente «útiles» ante problemas concretos (Newell, 1962). Por otra parte, cuando se

trataba de enfrentarse a problemas complejos del mundo real, los programas invariablemente fracasaban, sin que resultase claro cuáles eran las limitaciones en este sentido, así como si éstas dependían de factores cuantitativos a resolver con el transcurso del tiempo y el perfeccionamiento del hardware (la velocidad de los procesadores o la capacidad de los dispositivos de memoria), o si por el contrario existía algún techo cualitativo intrínseco para los problemas que podían resolver los computadores.

A principios de los años setenta, los investigadores británicos descubrieron la llamada «explosión combinatoria», e identificaron en ella la más grave cortapisa para el desarrollo e la IA (Lighthill, 1973, cit. en Boden, 1977; Aleksander, 1986; Nilsson, 1982; Stefik, *et al.*, 1983a). Se trata de un fenómeno simple, pero de una contundencia despiadada: a medida que aumenta la complejidad de un problema, la cantidad de posibles combinaciones de decisiones crece a un ritmo tal que hace completamente imposible su evaluación completa en un tiempo lo bastante breve para resultar eficaz. Por ejemplo, volviendo al conocido caso del ajedrez, si tomamos una posición cualquiera del desarrollo del juego, el promedio de jugadas lícitas que parten de ella es de unas 35. Esto implica que la evaluación exhaustiva de las posibles continuaciones a lo largo de tres jugadas exigiría examinar más de 1.800 millones de movimientos. Por supuesto, este valor es mucho mayor en ciertos momentos de una partida (después de las maniobras iniciales, por ejemplo); y un juego eficaz exige frecuentemente maniobras que tengan en cuenta una prospectiva de más de tres jugadas (Waltz, 1982). Y, desde luego, en problemas de mayor complejidad la explosión combinatoria tiene efectos aún más espectaculares. Pero el problema es siempre el mismo: la búsqueda exhaustiva hace que mucho antes de haberse alcanzado un mínimo nivel de eficacia, la capacidad de cómputo de la máquina quede desbordada.

Por tanto, si la IA ha de ser útil en problemas reales y complejos, resulta necesario «podar» el árbol de búsqueda. Esto significa emplear algún método de decisión en virtud del cual algunas ramas sean exploradas y otras no, reduciéndose selectivamente la amplitud del espacio de la tarea (Nilsson, 1982; Barr y Feigenbaum, 1982a).

Una forma obvia de reducir la búsqueda es eliminar una cierta parte de las alternativas (esto es, podar parte de las ramas) al azar (Waltz, 1982). En algunos problemas con muchas soluciones igualmente válidas esto puede suponer una pérdida pequeña de probabilidad de éxito (calculable a veces por métodos estadísticos). Por tanto, puede ser una solución útil en ocasiones. Sin embargo, resulta difícil sostener que programas basados en tales procedimientos puedan considerarse, aun en sentido metafórico, «inteligentes».

Volviendo al ejemplo clásico del jugador humano de ajedrez, cualquier psicólogo cognitivo puede dar unas cuantas razones de peso por las cuales es imposible que utilice una estrategia de búsqueda exhaustiva. Las limitaciones de capacidad bruta del procesador humano son claras a estos efectos. Pero, por otra parte, en absoluto parece que los jugadores humanos (al menos los expertos) se guíen por el azar. Esto es inmediatamente generalizable a una multitud de tareas y problemas muy diferentes del juego de ajedrez: invariablemente, el procesador humano se dirige certeramente al objetivo mediante estrategias que no parece que puedan calificarse ni de exhaustivas ni de aleatorias.

Pero existen otras estrategias cuyas posibilidades de éxito se aproximan a las de las primeras, sin que su consumo computacional resulte excesivo. Tales son los *métodos heurísticos*. Los heurísticos se pueden definir como reglas inexactas capaces de guiar los procesos de un agente cognitivo enfrentado a un problema con grandes posibilidades de éxito (Lenat, 1982; Pearl, 1984, y Negoita, 1985). Esas altas probabilidades de éxito se deben a las reglas heurísticas están fundamentadas sobre la experiencia. En una regla heurística se postula, por lo general, una relación empírica entre varios eventos, sin que esa relación tenga que poseer un carácter causal o necesario (Riedl, 1983).

La noción de heurístico es muy familiar para nosotros los psicólogos, hasta el punto de que supone un tema central de áreas muy significativas de nuestra disciplina, como la teoría de la decisión conductual (véase, por ejemplo, Tversky y Kahneman, 1982; Gambará y León, en prensa). Sin embargo, tanto su significación como su uso son diferentes en el contexto de la IA. Los métodos de búsqueda basados en heurísticos se han utilizado desde el mismo origen de la IA (Barr y Feigenbaum, 1982a), pero, a raíz de la problemática suscitada por la explosión combinatoria, su utilización se ha extendido hasta el punto de que hoy podemos considerarlos como un segundo rasgo característico de los programas de IA (Lenat y Brown, 1984; Pohl, 1977, y Newell y Simon, 1981).

Un ejemplo de regla heurística podría ser, siguiendo con la temática ajedrecística, el siguiente: «Si la partida está en fase inicial, es preferible efectuar movimientos que tiendan a cubrir un área grande del centro del tablero.» Es importante observar que, en primer lugar, se trata de una regla bastante imprecisa —sobre todo si se considera que en la informática tradicional el criterio de «precisión» es la absoluta ausencia de ambigüedad— y, en segundo lugar, que no siempre va a ser válida: en ciertas situaciones concretas (aunque no inventariables *a priori*), los movimientos contrarios pueden hacer desaconsejable aplicarla. Pero cualquier jugador mediano de ajedrez convendrá en que, en general, resulta útil seguirla, y probablemente él mismo tendrá que reconocer que lo hace. Sin embargo, resultaría difícilísimo fundamentarla en una teoría lógica del ajedrez o deducirla *formalmente* de las propias reglas del juego. En suma, tiene una elevada (aunque difícil de cuantificar) probabilidad de éxito, pero no se basa en un razonamiento lógico a partir de un conjunto bien definido de premisas.

Podemos poner otro ejemplo, esta vez con una temática psicosocial: «Si una persona habla en voz alta y con brusquedad, entonces seguramente es radical en sus actitudes y opiniones.» Muchos psicólogos relacionarían inmediata y acertadamente tal enunciado con temas como los estereotipos. Evidentemente, se trata de una regla que, como aquellos, carece de cualquier fundamento lógico-formal. No obstante, es interesante observar que esta regla nos permite tomar una decisión sobre cómo tratar a una persona desconocida, de la cual poseemos sólo información superficial, con razonables posibilidades de éxito (por ejemplo, mostrarnos inicialmente cautos en cuanto a las opiniones que enunciamos en su presencia). A falta de reglas de esta naturaleza nos veríamos abocados a demorar esa decisión hasta contar con muchos más elementos de juicio; pero, entre tanto, podemos haber perdido la oportunidad de modular la relación con esa persona de forma conveniente para nuestros objetivos. En definitiva, el heurístico es infor-

mal e incierto, pero adaptativo. Intuitiva e introspectivamente, parece que al menos buena parte de nuestro comportamiento inteligente podría formularse como aplicación de reglas heurísticas. Hoy día, muchos investigadores en IA consideran que el éxito y la eficacia de un programa dependen en buena medida de los heurísticos que emplee (Newell y Simon, 1981; Dehn y Shank, 1982; Hayes-Roth, *et al.*, 1983; Lenat, 1983; Lenat y Brown, 1984; Stefik, *et al.*, 1983; Pearl, 1984, y Waterman, 1986).

Durante la década de los setenta, el grueso de los investigadores se centran en la aplicación de los heurísticos para conseguir técnicas de búsqueda más eficaces, y en la implementación de formatos representacionales potentes y flexibles. Todo ello se aplica a problemas un tanto más específicos que en la década anterior (Gardner, 1982). Como resultado de ello, se producen avances importantes en campos como el análisis del lenguaje natural (Winograd, 1972; Schank, 1973, y Schank y Abelson, 1977), la demostración de teoremas (Kowalski y Kuchner, 1971; Chang y Lee, 1973, y Loveland, 1978), el reconocimiento visual (Rosenfeld y Kak, 1976, y Marr, 1978), etc.; y se desarrollan notablemente lenguajes y entornos de programación (las nuevas versiones del LISP, el Prolog...), así como conceptos y formalismos representacionales (las redes semánticas, los sistemas de producciones o los «frames»). Con todo, habrá que esperar al final de la década para que se establezca de forma generalizada el avance más significativo para el tema que nos ocupa: el cambio hacia problemas enormemente específicos y la idea clave, según la cual el poder de un programa depende, tanto o más que de los formalismos representacionales o los métodos de inferencia que utilice, de su posesión de un refinado *conocimiento* sobre el campo temático a manejar (Waterman, 1986). Ello nos sitúa ya en los sistemas expertos.

Como síntesis de esta pequeña revisión, podemos acotar el campo de la IA como la construcción de programas de computador destinados a la resolución de problemas complejos y de muy variada naturaleza mediante el empleo de métodos de computación «inteligentes», entendiendo por ello que no se basan ni en el azar ni en el análisis exhaustivo, sino en el manejo heurístico de relaciones muy sofisticadas entre elementos del espacio del problema.

LOS SISTEMAS EXPERTOS

Definición, filosofía y aplicaciones

Como sugiere su nombre, un sistema experto (SE) es un programa que trató de reproducir la actuación de un experto humano en un dominio de conocimiento altamente especializado.

Tratando de ser un poco más precisos, podemos considerar la definición aprobada por el comité de especialistas en Sistemas Expertos de la British Computer Society:

«... La incorporación en un ordenador de un componente basado en conocimiento que se obtiene a partir de la pericia de un experto, de forma tal que el sistema pueda dar CONSEJOS INTELIGENTES o tomar una DECISION INTELIGENTE acerca de una función de procesamiento.

Una característica adicional que es deseable, y que para muchos es fundamental es la capacidad del sistema para, bajo demanda, JUSTIFICAR SU PROPIA LINEA DE RAZONAMIENTO de una forma inmediatamente inteligible para el que pregunta.» (Tomado de Naylor, 1983, p. 13; mayúsculas del original.)

En esta definición, ciertamente, se encuentran algunos de los puntos fundamentales que caracterizan a los SE. No obstante, podría resultar algo oscura, y por ello vamos a intentar complementarla con un perfil más completo.

Ante todo, los SE son *programas de computador*. Todo programa de computador tiene como fin la resolución de un determinado tipo de problemas o tareas (aunque, si pensamos en ciertos contextos, el uso de los ordenadores se describiría mejor como una laboriosa búsqueda de problemas a los que aplicar la solución —el computador— que ya tenemos. Tal parece haber sido el caso, por ejemplo, en las fases tempranas de los computadores «domésticos» o de las aplicaciones educativas). Por tanto, para iniciar una caracterización de los SE podemos fijarnos en el tipo de problemas o tareas que tratan de resolver.

Los problemas para los que se diseñan los SE son problemas que normalmente requieren el concurso de un agente humano que posee un profundo conocimiento de los mismos. Por otra parte, no resultan reductibles a procedimientos efectivos (algoritmos), sino que exigen la toma de decisiones en condiciones de información insuficiente o imprecisa. Ejemplos típicos de tareas a las que se destinan SE son el diagnóstico médico, la detección de averías en mecanismos complejos, la predicción meteorológica, la determinación de la estructura de las moléculas, etc. Tomando el primer ejemplo, uno de los más famosos y veteranos sistemas expertos, MYCIN (Shortliffe, 1976 y Buchanan y Shortliffe, 1984), se dedica al diagnóstico de enfermedades infecciosas mediante la identificación del agente causante. La forma de operar del programa es muy similar a la de un especialista humano: típicamente, el sistema comienza por pedir al usuario una serie de datos generales sobre el paciente (sexo, edad...) y a continuación pasa a pedir información sobre los síntomas que presenta y los datos de que ya se disponga sobre el origen de su enfermedad. Cada una de las respuestas que el usuario humano va dando al programa influye en las ulteriores preguntas de éste. En ocasiones, el usuario proporcionará al programa respuestas inseguras, con una probabilidad estimada asociada. Puede ocurrir también que no exista respuesta conocida para algunas de las preguntas que plantee el sistema, sin que ello impida en absoluto el funcionamiento del mismo: simplemente probará otras vías. Cuando el sistema está en condiciones, ofrece al usuario un listado de los posibles organismos causantes del trastorno, con probabilidades asociadas, así como una o más posibles terapias para combatir a los antígenos que haya identificado como más probables. En el caso de ciertos medicamentos, el programa se cerciorará de que no existan contraindicaciones o incompatibilidades, en cuyo caso variará la prescripción. En cualquier momento del proceso, el usuario puede pedir al programa que justifique en qué se ha basado para hacer una determinada pregunta o para emitir un determinado juicio. Por ejemplo, si el sistema pregunta al usuario si un determinado análisis cuyo resultado se le ha comunicado se ha efectuado a partir de una muestra de sangre, de orina, o de

otro medio, aquel puede preguntar al sistema «¿por qué?». Este responderá, por ejemplo, que si el cultivo procede de sangre, entonces se puede aplicar una regla determinada que arroja tal o cual consecuencia respecto a la identificación del antígeno.

Así pues, los problemas que resuelven los SE exigen, en primer lugar, tomar decisiones basándose en una *información insuficiente*. Cualquier aficionado a la programación sabe que es muy fácil redactar un programa que nos vaya pidiendo datos para luego tomar decisiones, pero que es mucho más difícil conseguir un programa capaz de continuar funcionando en ausencia de una parte de esos datos. Pero, además, en los problemas que se plantea a los SE es completamente imposible predecir cuáles informaciones van a estar disponibles y cuáles no en cada caso (es decir, los datos van a ser «arbitrariamente insuficientes»). Ello hace inviable la resolución estrictamente algorítmica del problema.

En segundo lugar, los datos que estén disponibles van a tener *distintos grados de certeza*. Evidentemente, el programa no puede tratar del mismo modo un dato seguro que un dato con probabilidad, pongamos por caso, 0.7 de ser verdadero en un caso dado. Y el sistema tiene que ser capaz de enfrentarse con cualquier combinación de datos asociados a probabilidades, determinar cuándo puede llegar a alguna conclusión mínimamente fundamentada y, en definitiva, sacar de ellos lo que pueda en cada caso.

En tercer lugar, el sistema experto *se enfrenta a demandas variables*. Un programa de ajedrez, por potente, complejo, sofisticado e inteligente que sea, se enfrenta siempre en cierto modo al mismo problema, guiado por un objetivo fijo: el jaque mate. Sin embargo, el tipo de problemas a los que se enfrenta un sistema experto tienen una variabilidad mucho mayor, al menos en dos sentidos: de una parte, el funcionamiento con datos arbitrariamente insuficientes y probabilísticos supone ya de por sí un gran número de problemas computacionalmente diferentes. Es decir, desde un punto de vista formal, un mismo caso clínico, procesado con distintas cantidades de información y distintos grados de certeza no es «el mismo problema», por más que el objetivo sea siempre el diagnóstico y el tratamiento. Pero además, en los SE el usuario tiene normalmente la posibilidad de modular la prioridad que se dé a los distintos objetivos posibles. Por ejemplo, cabe pedir al programa que nos proporcione una terapia menos rápida, pero que no incluya un determinado medicamento; o podemos pedir únicamente un diagnóstico con un grado optimizado de certeza o por el contrario una lista que haga hincapié en la exhaustividad.

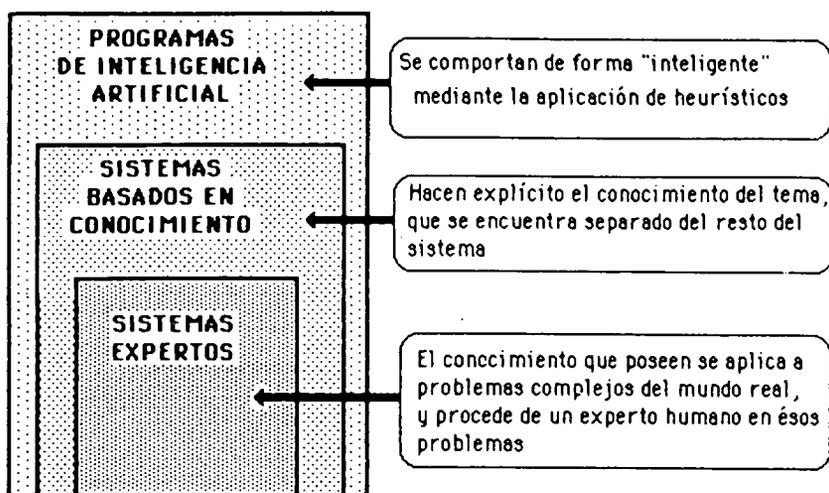
Estructura

Nótese que la caracterización de los sistemas expertos que hemos hecho hasta ahora es puramente funcional: nos hemos limitado a describir «qué es lo que tienen que hacer»; a qué problemas se enfrenta y qué condiciones han de cumplir en su funcionamiento. A continuación vamos a fijarnos en la organización computacional que permite el cumplimiento de esos cometidos.

Cuando se construye cualquier programa de computador se introducen en él diversos tipos de información. Por ejemplo, es necesario incluir tanto

conocimientos sobre cómo resolver el tipo de problemas para los que se construye el programa, como datos y procedimientos para abordar cada problema particular, variables locales, algoritmos y datos para la interacción usuario-programa, etc. En la mayoría de los programas, todos estos tipos de información se encuentran mezclados en el código, de forma que es prácticamente imposible distinguirlos o separarlos conceptualmente (Fernández, 1986; Stefik, *et al.*, 1983b). Pues bien, una parte importante de los programas actuales de IA se caracterizan precisamente porque en su construcción se encuentra perfectamente diferenciado el componente que incorpora el conocimiento relativo a la clase de tareas que el programa es capaz de realizar. Ese componente es la «Base de Conocimientos» del programa, y tales programas se denominan *Sistemas Basados en el Conocimiento*. Tal separación del conocimiento sobre la tarea del resto del programa facilita la consecución de dos importantes y deseables características: flexibilidad para adquirir y modificar los conocimientos, y posibilidad de acceso a los razonamientos y decisiones efectuados por el sistema (Fernández, 1986). Frecuentemente, en la literatura se identifican los sistemas expertos con los sistemas basados en conocimiento. Esto no es en la práctica demasiado inexacto; sin embargo, en rigor, es más correcto considerar a los primeros como un subconjunto de los segundos (Waterman, 1986), caracterizados por el hecho de que el conocimiento que contienen procede de especialistas en disciplinas específicas del mundo real. En la figura 2 se representa esta relación.

FIGURA 2



La relación de los sistemas expertos con los Sistemas Basados en Conocimiento y la IA en general (según Waterman, 1986).

La obtención de ese conocimiento a partir de un experto humano es, por sí sola, una tarea de enorme complejidad. Una parte fundamental del mismo es refractaria a la introspección, y para el propio experto resulta muy difícil formular explícitamente numerosas reglas, indicios, trucos, etc., que *de hecho* utiliza en su práctica cotidiana (Cuenca, 1986b). La elicitación de estos contenidos y su reducción a un formato implementable en un computador constituye el núcleo de una nueva disciplina: la *ingeniería del cono-*

cimiento. La ingeniería del conocimiento se define a menudo, simplemente, como «la construcción de sistemas expertos» o «el diseño de sistemas basados en conocimiento» (Waterman, 1986 y Cuenca, 1986a). Sin embargo, el papel del ingeniero del conocimiento resulta perfectamente diferenciable del de otros profesionales que intervienen en la construcción de sistemas expertos (como, por ejemplo, los programadores). Le corresponden, como hemos indicado, dos tipos fundamentales de tareas: por una parte, conseguir que los expertos humanos hagan explícito su conocimiento (lo cual implica a menudo una larga y laboriosa cadena de entrevistas, formulaciones provisionales, etc.) y, por otra, una serie de decisiones en cuanto al formato representacional, método de inferencia y otros aspectos de la estructura del sistema experto que lo hagan adecuado para la naturaleza del tema (en cooperación con analistas, programadores, etc.). El principal resultado de todo ello es una base de conocimientos que compendia en un formato computable el conocimiento extraído.

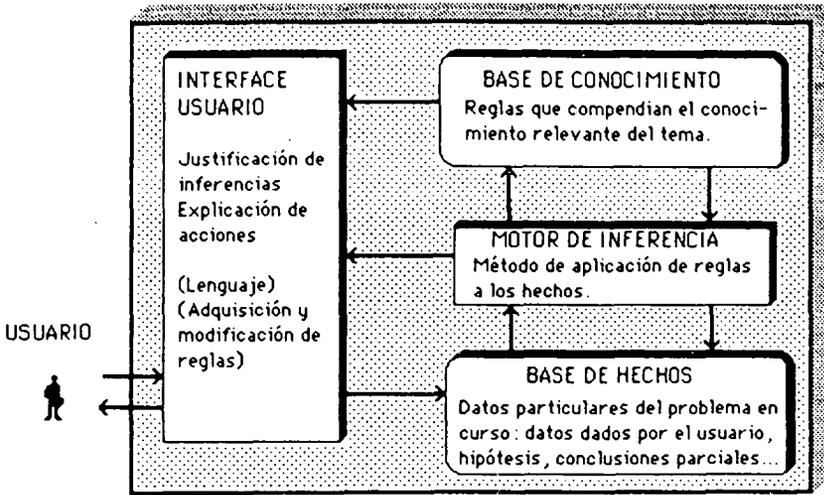
Es importante señalar que el estado de desarrollo de la ingeniería del conocimiento hoy en día es aún muy precario. Se pueden diferenciar dos enfoques: en primer lugar, las tareas antedichas pueden acometerse conforme hemos indicado: mediante entrevistas, sesiones de trabajo conjuntas con los expertos, etc. Pero existe también la posibilidad de automatizarlas total o parcialmente: un sistema experto puede incorporar funciones de extracción de conocimientos mediante interacción con especialistas humanos. De hecho, se han realizado ya interesantes experiencias en esta dirección (véase, por ejemplo, Davis y Lenat, 1982), que promete importantes desarrollos a corto y medio plazo (Cuenca, 1986a y Barr y Feigenbaum, 1982). Obsérvese que estas cuestiones son, ya a primera vista, muy interesantes para la PC. Por ejemplo, buena parte de las tareas del ingeniero del conocimiento se asemejan llamativamente al método clínico piagetiano (Piaget, 1981), en la medida en que se trata de obtener mediante el diálogo unos contenidos mentales de un sujeto humano, a partir de los cuales se intenta inferir una estructura formalizable. Vamos, sin embargo, a posponer la discusión de estos aspectos y a centrarnos en la constitución interna de un sistema experto.

En la figura 3 se representa un esquema idealizado de las partes de un sistema experto. Hay que señalar que dicho esquema entraña un grado notable de simplificación. Por una parte, no es igualmente apropiado para todo sistema experto —por ejemplo, no representaría adecuadamente a los sistemas expertos basados en arquitecturas de «pizarra» (Lenat, 1984; Hayes-Roth, *et al.*, 1983)— y, por otra parte, se omiten módulos fundamentales de algunos sistemas particulares (o bien se los da por incluidos en otros componentes). Sin embargo, el esquema representa de forma sencilla la arquitectura característica de la mayoría de los SE, cuyos componentes vamos a examinar brevemente por separado.

Base de conocimiento

El primer y más característico componente estructural de un sistema experto va a ser, pues, la *base de conocimiento*, que contiene el conocimiento específico del área en que el sistema es experto. Las técnicas representacionales empleadas varían de unos sistemas a otros, pero la mayoría se pueden

FIGURA 3



Componentes principales de un sistema experto idealizado (adaptado de Fernández, 1986).

incluir en dos grandes tipos: *representación basada en reglas* y *representación basada en redes semánticas*. Las primeras son las más frecuentes y quizá las más típicas de los SE; por ello, les vamos a dedicar mayor atención. Las segundas dan como resultado bases de conocimiento altamente estructuradas, muy adecuadas para dominios de conocimiento susceptibles de ser expresados como taxonomías (Waterman, 1986). En este tipo se pueden incluir formatos representacionales tan clásicos como las redes de Quillian (1968) o los «frames» de Minsky (1975).

La base de conocimiento de los sistemas basados en reglas consta de un determinado número (por lo general, entre unas decenas y varios centenares) de reglas heurísticas del tipo «SI... ENTONCES...». Se trata de conocimiento *procedimental* (Fernández, 1986): las reglas expresan operaciones (predicaciones, atribuciones, evaluaciones, etc.) a efectuar en circunstancias determinadas. Una cuestión fundamental es el formalismo concreto a que se adopte para introducirlas en el sistema. A este respecto, la opción casi general son los sistemas de producciones.

Los sistemas de producciones son un formalismo computacional descrito por primera vez por Post (1943, cit. en Fernández, 1986), y muy populares entre los psicólogos cognitivos debido a su profusa utilización en numerosos trabajos de gran trascendencia en nuestra ciencia (vg. Newell y Simon, 1972 y Anderson, 1976, 1983). Como es sabido, constan de un término antecedente (condición) que expresa la situación en la cual la regla será de aplicación, en términos de la presencia o ausencia de determinados hechos, unidas por conectivas lógicas (normalmente conjunción y disyunción); y de un término consecuente, que expresa las acciones a acometer en caso de que la regla resulte de aplicación (es decir, si se verifica la situación expresada en la condición). Las reglas de producción poseen ventajas importantes en cuanto formato para representar conocimiento: en primer lugar, permiten expresar el conocimiento de un dominio en términos próximos a como un experto humano puede comunicarlo en lenguaje natural.

En segundo lugar, cada regla puede representar un aspecto relativamente «completo» del conocimiento, puesto que en el antecedente se determina inequívocamente la situación a que se refiere la regla. En tercer lugar, una regla de producción no tiene que hacer referencia a otras, lo cual permite poner, quitar y cambiar reglas sin necesidad de revisar exhaustivamente cada vez todo el sistema, al contrario que ocurre con otros formalismos en los que por ejemplo una subrutina es llamada desde numerosos puntos del sistema, y modificarla puede suponer un desajuste de gran alcance (Davis y Lenat, 1982). Al mismo tiempo, las reglas de producción son fácilmente expresables en lenguajes de alto nivel (en el caso del LISP, por ejemplo, la traducción resulta muy directa). Por tanto, resultan también adecuadas desde el punto de vista del programador. (Para una discusión en profundidad de los sistemas de producción, sus ventajas y sus inconvenientes, véase, por ejemplo, Davis y King, 1977 o Newell, 1973b.)

Por ejemplo, una regla integrante de la base de conocimientos de un imaginario sistema experto en mecánica del automóvil podría ser:

SI	dos o menos componentes del sistema eléctrico no funcionan
	y los restantes componentes del sistema eléctrico funcionan
	y el vehículo tiene menos de 100.000 km.
ENTONCES	(0.6) la localización de la avería es en los fusibles

El ejemplo sirve para ilustrar varias cuestiones fundamentales que es necesario resolver al diseñar la base de conocimientos: debe incorporar una serie de conceptos o «primitivas semánticas» (en el ejemplo anterior, «vehículo», «sistema eléctrico», etc.), así como una serie de predicados («funcionan»), atributos («localización») y valores («en los fusibles») que permitan describir las relaciones entre esos conceptos y establecer las evaluaciones pertinentes (Barstow *et al.*, 1983). Otro problema importante es la incorporación de algún método para la descripción del grado de certidumbre de las reglas (en el ejemplo, el «factor de certidumbre» 0.6). Ello es necesario no sólo porque a las reglas particulares se les atribuyan diferentes grados de confianza (a partir del juicio de los expertos humanos), sino también porque la validez de una conclusión va a interpretarse en función de los factores de certidumbre de todas las reglas aplicadas para alcanzarla (Naylor, 1983 y Fernández, 1986).

En definitiva, la base de conocimiento es una estructura simbólica que representa el conocimiento relativo a un determinado dominio temático, es decir, sus elementos conceptuales y las relaciones relevantes entre ellos.

Base de hechos

En cualquier proceso de resolución de un problema, además del conocimiento sobre el mismo, es necesario disponer de información, a veces extensa y compleja, sobre los pormenores del problema particular que en un momento dado se está abordando. Esa es la finalidad de lo que en los sistemas expertos se suele denominar «base de hechos»: contener los datos particulares del caso, así como diversos tipos de datos sobre el estado del sistema. Por ejemplo, en la base de hechos se van acumulando las conclu-

siones parciales a las que se va llegando, su grado de certeza, las reglas que se han ido aplicando, los datos que proporciona el usuario acerca del caso, etc.

En realidad, la base de hechos difiere bastante poco de cualquier base de datos convencional. A diferencia de la Base de Conocimiento, se trata simplemente de un conjunto de datos organizado, cuya principal cualidad debe ser una adecuada accesibilidad para el sistema que actúa sobre ella (Waterman y Hayes-Roth, 1983); y su estructura sólo se comprende desde el punto de vista de esa actuación. Dicho de otro modo, no se trata como en la base de conocimiento de conocimiento altamente refinado, sofisticado e integrado, sino simplemente de información estructurada de forma eficaz desde un punto de vista operativo.

Motor de inferencia

Con lo descrito hasta ahora, tenemos un componente que integra el conocimiento sobre el tema y otro que almacena la información relevante para los casos concretos. Pero se trata, desde luego, de componentes relativamente «pasivos». Se requiere un tercer componente o módulo que se ocupe de aplicar el conocimiento a los hechos: el motor de inferencia.

Más específicamente, el motor de inferencia supone la introducción en el sistema de un método inferencial que se ocupa de decidir qué reglas procede aplicar en cada momento y evaluar el resultado de su aplicación. Esto implica, por supuesto, todo un proceso de selección de reglas o subestructuras de la base de conocimientos, toma de datos de la base de hechos e incorporación de conclusiones, hipótesis y otros elementos a la misma, que es lo que a fin de cuentas permite avanzar en la resolución del problema.

En realidad, es difícil describir un motor de inferencias como módulo independiente, ya que a menudo, en la práctica, no lo es. Frecuentemente, por ejemplo, los lenguajes de más alto nivel para la construcción de sistemas expertos llevan ya implícito un mecanismo inferencial (Barstow *et al.*, 1983); en cambio, si se utilizan lenguajes de nivel inferior, es necesario diseñarlo como programa independiente.

Lo que sí tiene entidad propia e independiente es el método de inferencia que se implemente (Verdejo, 1986 y Waterman y Hayes-Roth, 1983). Y aquí hay dos posibilidades fundamentales: Un motor de inferencias basado en el método conocido como «encadenamiento hacia atrás» (*backward chaining*), comienza por establecer una hipótesis a comprobar, y a continuación buscará en la base de conocimientos la regla o reglas que sirvan para tal comprobación (es decir, aquellas que contienen en su consecuente el enunciado que se quiere demostrar). Entonces, si se constata que los antecedentes o «condiciones» de esas reglas son válidos, la hipótesis también lo será, con lo que dichos antecedentes pasan a ser tratados como nuevos objetivos a comprobar. Este proceso continúa hasta que se alcanza información que conste como válida en la base de hechos (por ejemplo, datos del planteamiento del problema en curso introducidos por el usuario). La hipótesis inicial queda así justificada por esa información, a través de una cadena de reglas verificadas. Este método se suele llamar «guiado por objetivos» (Verdejo, 1986), ya que parte de lo que se quiere demostrar y le

busca un fundamento en la base de conocimientos. Un segundo método general es el encadenamiento hacia adelante (*forward chaining*), en el cual el consecuente de la regla se ejecuta si se cumple el antecedente, y esa ejecución puede suponer evaluar una nueva regla (en cierto modo, es una forma más «literal» de interpretar el contenido procedural de un sistema de producciones).

La «instanciación» concreta de esos métodos generales en los distintos sistemas es muy variable. Pero además, un motor de inferencias debe solucionar otros problemas. Por ejemplo, es necesario que incorpore algún método para resolver «conflictos» entre reglas, es decir, para determinar cuál aplicar primero de entre un conjunto de reglas que en un momento dado satisfagan por igual los criterios en curso. De la existencia de heurísticos potentes capaces de resolver tales conflictos depende en gran medida la eficacia general del comportamiento del sistema.

Otros componentes

Con los módulos descritos hasta ahora, tenemos un esbozo de las partes esenciales que capacitan al sistema para tomar decisiones inteligentes en un campo temático específico. Como ya hemos señalado, tal esbozo no es sino una simplificación que reflejaría un sistema experto idealizado. En realidad, hay muchos otros aspectos que varían mucho más de unos sistemas a otros, tanto en cuanto al modo en que se resuelven computacionalmente como respecto al énfasis que reciben. Algunos de ellos son tan importantes (véase la definición en el apartado 3.1) que no podemos abandonar esta descripción sin al menos tratarlos someramente.

Lenguaje natural

La IA ha tropezado a menudo con un hecho que no por conocido deja de ser paradójico: muchas de las actividades cognitivas que los sujetos humanos aparentemente efectúan con más facilidad resultan ser las más difíciles de implementar en un sistema digital (García Albea, 1986a; considérese, por ejemplo, el reconocimiento visual de objetos: Marr, 1986a, b). El lenguaje natural, que todas las personas manejan con soltura (a diferencia de las matemáticas o la lógica), presenta una enorme dificultad para su estudio científico en general. Recientemente, Belinchón (1987) se refería a él como «algo así como un monstruo de mil cabezas que parece simple y familiar al hombre de la calle pero que resulta inabarcable para el psicólogo científico de hoy» (p. 210). De forma semejante, también la IA ha encontrado en el lenguaje natural uno de los problemas más difíciles de abordar con sus métodos y conceptos. No es éste el lugar adecuado para discutir la problemática computacional asociada al estudio del lenguaje (a este respecto, véase por ejemplo Winograd, 1983; Barr y Feigenbaum, 1982a, pp. 225 y ss.; o, desde una perspectiva práctica, Schneiderman, 1986).

Sin embargo, conviene señalar un aspecto del lenguaje que resulta particularmente significativo desde el punto de vista de los sistemas expertos: desde etapas muy tempranas de la IA se ha observado que un sistema de

procesamiento no puede manejar satisfactoriamente un lenguaje natural (vg., intervenir en una conversación) si le es totalmente ajeno el dominio de referencia (Quillian, 1968; Schank y Abelson, 1977). Es decir, el lenguaje presupone un cierto conocimiento de aquello de lo que se habla. Nuestras emisiones dependen decisivamente de un dominio de conocimientos tácitamente compartido por nuestro interlocutor y conlleva en este sentido un complejo cúmulo de convenios y pactos (Rivière, 1984). Los programas que han intentado incorporar lenguaje natural sin conocimiento, como *Eliza* (Weizenbaum, 1976), *Parry* (Colby, 1981) o *Racter* (Chamberlain, 1984) quedan en el dominio de lo anecdótico.

Uno de los más interesantes análisis de este problema lo constituye la investigación en torno al programa SHRDLU (Winograd, 1972). Dicho programa es capaz de conversar de manera más que aceptable con un sujeto humano, con tal de que el tema se limite a un universo restringido: el «mundo de bloques». La clave consiste en que el programa posee un modelo interno («conocimiento») de ese mundo, y es precisamente eso lo que le permite no sólo mantener una conversación, sino también actuar siguiendo instrucciones, aprender nuevos conceptos a partir de los que ya posee, etc.

El mundo de bloques es, desde luego, muy limitado, en el sentido de que puede describirse de forma completa y precisa mediante un conjunto relativamente reducido de enunciados. Evidentemente, no es ése el caso del mundo al que se refiere el lenguaje natural empleado por las personas. Por tanto, parece que para que sea posible una comunicación fluida entre un hombre no técnico y un programa es preciso que este último incorpore una ingente cantidad de conocimientos sobre el mundo en general, las convenciones del lenguaje, etc.

Así pues, ¿es imposible que un sistema experto, que posee un conocimiento muy específico y restringido, se comunique en lenguaje natural con el usuario? Ciertamente, existe un acuerdo generalizado en cuanto a que esto resulta muy deseable, si no necesario. Desde un punto de vista práctico, las aplicaciones más obvias de los SE suponen utilizarlos como ayudantes o colaboradores de los expertos humanos. Si el sistema requiere conocimientos técnicos para interactuar con él se corre el peligro de que termine por crear más problemas al profesional de los que le resuelve. Pero además, las investigaciones en esa dirección son terreno abonado para potenciales hallazgos de importancia para la ciencia cognitiva, y constituyen un campo de indudable interés teórico.

De hecho, muchos sistemas expertos actuales resuelven el problema de forma bastante satisfactoria a efectos prácticos: el usuario no tiene necesidad de aprender complicadas listas de comandos y normas sintácticas para manejarlos, sino que puede servirse del lenguaje natural. Y ello sin utilizar recursos computacionales diferentes de los comunes en otras áreas de la IA. En realidad, la capacidad lingüística de los SE descansa en gran medida en dos puntos: en primer lugar, el sistema posee de hecho conocimientos sobre su dominio temático, que no es sino un «mundo restringido» (aunque no tanto, por supuesto, como el mundo de bloques). El contenido de la base de conocimientos sirve, pues, como referente al lenguaje (Naylor, 1983). En segundo lugar, es frecuente que las actividades en las que un ser humano puede ser experto posean una «jerga» o *argot* peculiar, constituida por un conjunto de términos y construcciones mucho más reducido y es-

tereotipado de lo que podría parecer a primera vista (pensemos, por ejemplo, en el lenguaje profesional de los oftalmólogos, de los especialistas de *marketing* o de los técnicos en modificación de conducta). La combinación de estos dos hechos hace posible implementar en los sistemas expertos un subconjunto específico del lenguaje natural que les permite comunicarse con el usuario en términos cómodos y familiares para éste.

Adquisición y modificación del conocimiento

Como señalan Buchanan y Feigenbaum (1982), la adquisición y mantenimiento de grandes bases de conocimiento es un problema crucial para los sistemas basados en conocimiento. Ya hemos comentado algunas de las dificultades inherentes a la ingeniería del conocimiento, en lo que se refiere a simplemente obtener conocimiento de los expertos en un formato explícito. Pero, desde luego, el problema no termina ahí. Hay al menos dos aspectos más que debemos mencionar.

En primer lugar, dado el estado aún incipiente y un tanto «artesanal» de la ingeniería del conocimiento como disciplina, el resultado que genera suele ser una «tentativa» bastante aproximativa de compendiar el conocimiento del tema. Es un hecho generalizado que las bases de conocimientos requieren importantes modificaciones a partir de su construcción original (Buchanan *et al.*, 1983). Esto resulta obvio si tenemos en cuenta que los SE resuelven problemas complejos y sutiles, y, por tanto, una valoración firme de la adecuación de las reglas que maneja exige a menudo una larga experimentación con el sistema enfrentándose a casos reales. Normalmente, una primera versión exige modificar muchas reglas que producen comportamientos inadecuados. Por otra parte, en campos de rápido desarrollo, puede ser necesario incorporar conocimientos nuevos al sistema. En general, por tanto, es de máximo interés que las bases de conocimientos resulten flexibles y fácilmente modificables.

Actualmente hay varias tendencias dentro del campo que pueden ofrecer respuesta a estos problemas. En primer lugar, se han diseñado un buen número de «herramientas de construcción de sistemas expertos» (*expert system building tools* o simplemente *tools*), esto es, lenguajes de muy alto nivel dirigidos especialmente a la creación de sistemas expertos. Algunos de ellos permiten que una persona con escasos conocimientos técnicos interactúe con el sistema para crear la base de conocimientos. En ocasiones, se han construido complejos módulos que facilitan la comunicación con el usuario en esta tarea, incorporando potentes utilidades de «desparasitación» (*debugging*), y edición. Tal es, por ejemplo, el caso de *Teiresias* (Davis, 1976; Davis y Lenat, 1982), que actúa como una especie de ingeniero del conocimiento automatizado al permitir que el propio experto realice personalmente gran número de tareas en la puesta a punto y ajuste del contenido de la base de conocimientos sin necesidad de prestar atención a las cuestiones técnicas del sistema. Estos instrumentos poseen un interés intrínseco: por ejemplo, el diseño de *Teiresias* ha llevado a analizar en profundidad el problema del metaconocimiento (Davis y Lenat, *op. cit.*).

La estrategia aplicada para diseñar lenguajes o entornos de programación de sistemas expertos consiste a veces en partir de un sistema experto ter-

minado, y eliminar los contenidos específicos del tema para el que fue ideado en un principio. El resultado es lo que se denomina un «sistema vacío» o *shell* (concha, cascarón): un formato representacional, un motor de inferencia, etc., todo ello sin contenido particular. Tal sistema vacío sirve como base para la construcción de nuevos sistemas expertos, que aprovechan buena parte de las características formales del originario. Ese proceso se ha efectuado, por ejemplo, con el sistema MYCIN, anteriormente mencionado. Quitándole sus conocimientos sobre diagnóstico de enfermedades infecciosas, se obtuvo EMYCIN («Essential MYCIN») el cual ha servido de base para la construcción de sistemas expertos en temas muy diferentes, como PUFF, que se ocupa de enfermedades pulmonares, o SACON, destinado a problemas de ingeniería de estructuras (cfr. Barr y Feigenbaum, 1982b).

Un segundo aspecto fundamental relativo a la adquisición y modificación del conocimiento es la investigación sobre aprendizaje automático (*machine learning*). En la actualidad existe un amplio abanico de líneas de investigación en busca de métodos que permitan a un sistema adquirir por sí mismo nuevos conocimientos (véase Michalski, Carbonell y Mitchell, 1983; Mitchell, Carbonell y Michalsky, 1986, y Cohen y Feigenbaum, 1982). La implementación de tales métodos en los SE puede permitir que, por ejemplo, no sea necesario introducir reglas de manera expresa, sino que baste con proporcionarle casos o datos de los que el mismo pueda inducir las (Gardner, 1982; Simon, Langley y Bradshaw, 1981; Valiant, 1986, y Quinlan, 1979). En los casos en que, como se ha señalado frecuentemente, los expertos humanos no consiguen hacer explícitas las reglas en que se basan, la utilidad de los métodos de inducción automática abre posibilidades fascinantes, que trascienden en muchos casos el campo de la informática. El solo hecho de extraer regularidades a partir de conocimiento informal y tácito posee un interés intrínseco para la comprensión de la naturaleza del conocimiento.

El desarrollo de la investigación sobre sistemas expertos

A pesar de que, como ya hemos señalado, la generalización de los sistemas expertos se produce en la comunidad de la IA a mediados de los años setenta, en realidad las ideas clave venían siendo aplicadas en proyectos puntuales desde bastante antes. Siguiendo a Cuenca (1984), podemos establecer cuatro fases, en la evolución histórica de los sistemas expertos.

Entre los años 1965 y 1970 se construyeron los primeros sistemas explícitamente denominados «expertos», y ya fundados en la estructura que hemos descrito. El primero y más conocido de ellos es *Dendral* (Buchanan, *et al.*, 1969), desarrollado en la Universidad de Stanford, y destinado al análisis de la estructura molecular de compuestos químicos a partir de datos de espectrografía. Versiones mejoradas de *Dendral* continúan aplicándose en la actualidad en el trabajo práctico de los laboratorios químicos y farmacéuticos.

Entre 1970 y 1977 se desarrollan los sistemas más conocidos actualmente (Fernández, 1986). Entre ellos, cabe destacar a PROSPECTOR, sistema dedicado a la investigación geológica, y responsable de uno de los más es-

pectaculares éxitos prácticos de los sistemas expertos, al descubrir un yacimiento de molibdeno de gran valor económico. En este período se construye también MYCIN, el programa de diagnóstico médico de enfermedades infecciosas, del que ya hemos hablado.

A partir del año 1977, el reconocimiento de la importancia de los sistemas expertos en el contexto de la IA comienza a ser general, y se entra en los que Cuenca (1984) denomina una etapa de experimentación. En los años inmediatos, la investigación no se limita a construir nuevos sistemas expertos para nuevos fines, sino que además se desarrollan ideas que refinan cualitativamente las posibilidades de los mismos, como los *shells* o sistemas vacíos, y otras herramientas de programación. Se construyen también sistemas complementarios para facilitar la modificación de las bases de conocimientos y la interacción usuario-sistema, como el ya mencionado *Teiresias*. Por otra parte, aparecen sistemas cuyo interés está más en los avances conceptuales que implementan que en sus aplicaciones estrictamente prácticas. Tal es el caso de EURISKO (Lenat, 1983, 1984) que posee heurísticos generales aplicables a tareas tan diferentes como los juegos de salón y el diseño de microcircuitos.

Desde 1981 comienza una etapa de industrialización, con la formación de la primera compañía comercial (Teknowledge) destinada a la construcción de sistemas expertos y herramientas para su programación (Cuenca, *op. cit.*, y Fernández, 1986). En la actualidad han surgido muchas otras, y los sistemas expertos ocupan un lugar preferente en el mercado mundial del *software*.

DISCUSIÓN

Hasta aquí hemos esbozado una breve revisión de la evolución que se ha producido en el seno de la IA respecto a la manera en que se ha abordado el problema de describir y modelar conductas «inteligentes» con ayuda de computadores; y nos hemos centrado en la significación, la arquitectura y el desarrollo de los sistemas expertos, ya que constituyen, a esos efectos, una de las alternativas más prometedoras de la IA actual. Ahora, para concluir, discutiremos brevemente las implicaciones que estos desarrollos de la IA pueden tener en el contexto de la PC, puesto que, como ya hemos señalado, si bien tanto las perspectivas como los propósitos iniciales de ambas disciplinas son diferentes, lo cierto es que históricamente ha existido una fuerte interacción entre la IA y el PC (Newell, 1970, 1973a, y Seoane, 1979).

La relación psicología cognitiva-inteligencia artificial

La IA y la PC, en primer lugar, comparten un amplio *vocabulario conceptual*. La PPI, bloque mayoritario de la PC en la actualidad, se asienta en el supuesto de la analogía del ordenador. Como resultado de ello, numerosos términos del lenguaje conceptual informático son tomados por la PPI (proceso, algoritmo, programa, nodo, sistema de control, procesamiento paralelo/secuencial...). De forma similar, los especialistas en IA se ven

obligados a utilizar constantemente nociones de origen psicológico (inteligencia, exploración, percepción, *chunk*, etc.).

En segundo lugar, la adopción de la *metodología de simulación* por los investigadores en PC ha acercado a ésta aún más a la IA: el «producto» de las investigaciones psicológicas puede ahora ser un programa de computador, además (o en vez) de las tradicionales formulaciones verbales o matemáticas de teorías y modelos. Ello ha hecho que algunas teorías, siendo netamente «psicológicas» en la medida en que sus autores intentan explícitamente modelar el comportamiento específicamente humano, hayan sido asumidas por la IA como propias y sean objeto habitual de estudio y análisis por los investigadores de ésta. Tal es el caso, por ejemplo, de la teoría ACT de Anderson (1976, 1983), a la que se dedican capítulos enteros en los manuales de IA (cfr. Cohen y Feigenbaum, 1982). Evidentemente, lo mismo ocurre en sentido inverso (vg. la teoría de la solución de problemas de Newell y Simon, 1972).

En tercer lugar, un nuevo marco general se ha desarrollado a partir de los primeros años setenta: *la ciencia cognitiva*. La ciencia cognitiva asume como objeto de estudio todos aquellos sistemas cuyo comportamiento se caracteriza fundamentalmente por propiedades que se pueden describir y explicar adecuadamente en términos de manipulación de símbolos (Newell, 1980), propiedades de las cuales es imposible dar cuenta mediante descripciones o explicaciones más moleculares (Putnam, 1973; Newell, 1982; Pylyshyn, 1980, 1984, y Haugeland, 1981a). Así pues, se define una categoría de sistemas que se pueden estudiar a partir de un mismo nivel de descripción: el plano cognitivo (Rivière, 1986). Tal clase de sistemas abarca tanto el aparato cognitivo humano como los computadores digitales, los lenguajes formales y naturales, los autómatas, etc.; y las disciplinas tradicionalmente asociadas al estudio de cada uno de ellos (Lingüística, PC, IA, Lógica, ciertas ramas de las Matemáticas como la Teoría de Autómatas Formales, etc.) serían por tanto, en parte, asimilables a la ciencia cognitiva. Se supone que, puesto que todas ellas tienen problemas comunes, pueden asimismo compartir soluciones y aportaciones. Entre estas últimas, cabría destacar las importantes contribuciones de algunos filósofos a la ciencia cognitiva, los cuales vienen desempeñando una función integradora y sistematizadora de los desarrollos más punteros de ésta (Fodor, 1966, 1980; Putnam, 1973; Dennett, 1981; Dretske, 1981; Kyburg, 1983, y Sayre, 1986). Función que resulta crucial para guiar y coordinar los esfuerzos de una ciencia tan acusadamente interdisciplinar.

Entre las aportaciones de autores como los anteriormente citados, algunas son extremadamente significativas para la PPI, ya que ponen de manifiesto cuestiones que, siendo de vital importancia, tienden a ser obviadas en la práctica de la investigación. Por poner un ejemplo llamativo de algunos de estos trabajos parece desprenderse la inquietante conclusión de que la PPI podría estar incurriendo en errores semejantes a los que criticaba a la psicología conductista. El conductismo se fundamenta en un reduccionismo asociacionista: todo el comportamiento puede explicarse en términos de asociación entre los estímulos del medio y las respuestas del organismo. Sin embargo, en la práctica, los conductistas no son consecuentes con tal postulado, e «impregnan» sus explicaciones de contenidos cognitivos (para una minuciosa justificación de esta afirmación véase Pylyshyn,

1984). Con ello, el término «estímulo» queda cargado con un lastre atributivo muy superior a lo que puede soportar su desarrollo teórico por parte de los conductistas, el cual resulta particularmente insuficiente respecto a la relación entre el estímulo proximal y el estímulo distal (Fodor, 1983, 1985).

Curiosamente, el problema de la PPI con el concepto de «información» podría ser sometido a una crítica análoga (Sayre, 1986). El psicólogo del PI supone que, en último término, la actividad cognitiva es *reducible* a términos de información (De Vega, 1987). Tal presupuesto nos sitúa entre varias alternativas: podríamos, en primer lugar, tomar el término «información» en su sentido común o vulgar: la información sería entonces algo no muy bien definido que parece tener que ver con la «novedad» de un determinado «conocimiento» y que se transmite mediante enunciados verbales, actos, objetos como los periódicos o las señales de tráfico, etc. En este caso, desde luego, habría que admitir que toda la PPI se asentaría sobre una noción no científica. Podemos también tomar «información» en el sentido técnico de la teoría de la comunicación de Shannon, que la define como «reducción cuantificable de incertidumbre». Tenemos entonces ni más ni menos que una magnitud definida como función de los estados posibles de un sistema («fuente») y sus respectivas probabilidades. Esta segunda opción nos lleva a la necesidad de especificar teóricamente cómo se relaciona esa magnitud con los conceptos fundamentales que de hecho manejamos en nuestras explicaciones del comportamiento (como, por ejemplo, «inferencia», «contenido semántico», «conocimiento», «esquema», etc.). Tal especificación no es fácil ni siquiera para los investigadores de la IA, los cuales, a diferencia de los psicólogos, pueden examinar su objeto de estudio (los programas) a cualquier nivel molar o molecular de análisis. Una tercera opción sería tratar de elaborar un nuevo concepto teórico bien definido de «información» para nuestros propios fines.

En la segunda opción se enmarcaría, por ejemplo, la teoría de la percepción de Sayre (1986); y en la tercera encontramos el notable trabajo de Dretske (1981). Sin embargo, la mayor parte de la PPI no suele hacer explícito el concepto de «información» que maneja, y la actitud más generalizada es un «reduccionismo informacional» que, como el reduccionismo asociacionista, tampoco se aplica consecuentemente: *la noción de información queda «impregnada» de un «contenido intencional»* (en el sentido de Dennett, 1981, 1983), pero su desarrollo teórico por parte de la PPI es insuficiente para dar cuenta de esa carga.

Es muy significativo que este problema se presente igualmente en la IA (McDermott, 1976): los programas son sistemas capaces de preservar las propiedades semánticas (como la «verdad») de las estructuras simbólicas que manejan y transforman; pero dichas propiedades son extrínsecas a ellos (en sí mismos son sólo «sistemas de significantes que no significan nada»), y dependen de la interpretación que el investigador haga de ellos (Haugeland, 1981a). Y aquí se vuelve a dar, por parte de éste, la atribución intencional. La ciencia cognitiva, sin embargo, apunta actualmente hacia vías muy prometedoras de solución. En particular, el análisis explícito de la noción de «sistema intencional» (Dennett, 1981, y Haugeland, 1981a) parece especialmente significativo al menos a efectos de plantear adecuadamente el problema.

Al margen de estas cuestiones, lo cierto es que numerosos teóricos de la PPI, al adoptar el supuesto de los «sistemas simbólicos físicos» (Newell, 1980), se han alineado explícitamente dentro de la ciencia cognitiva. Tal situación ha servido para «formalizar la relación» que de hecho ya existía entre la IA y la PPI. Y al mismo tiempo ha sido objeto de debate interno en la psicología. Una parte de los psicólogos cognitivos no consideran adecuadas, desde un punto de vista psicológico, las posiciones adoptadas por algunos autores, según las cuales, por ejemplo, el hecho de que un programa ejecute de forma eficaz una función cognitiva dada es condición suficiente para —dadas unas mínimas restricciones formales— considerarlo una potencial teoría psicológica de dicha función (véase, por ejemplo, Pylyshyn, 1984). Como es sabido, planteamientos de esta naturaleza introducen problemas importantes (véase, por ejemplo, Rivière, 1986, 1987; Carello *et al.*, 1984). Sin embargo, desde el punto de vista de otros autores, si la alternativa es la consideración de lo computacional como una mera metáfora, con todo el peligro de imprecisión, incoherencia teórica e irrelevancia que ello supone (vg. Kolars y Smythe, 1984), entonces la adopción en sentido fuerte de los principios de la ciencia cognitiva podía ser ventajosa (véase, por ejemplo, García-Albea, 1986b).

En todo caso, y cualesquiera que sean las posiciones teóricas que se suscriban, es evidente que la relación entre la PPI y la IA ha sido y es muy estrecha. Como resultado, es difícil imaginar cómo sería la PPI en la actualidad sin numerosas aportaciones teórico-metodológicas de la IA (vg. sistemas de producciones, redes semánticas, etc.). Y tales aportaciones no se vierten únicamente en la PPI, sino también en otros enfoques de la PC, como el piagetismo (vg. Boden, 1981). Consiguientemente, es razonable plantearse la cuestión de qué alcance puedan tener en la PC los cambios y desarrollos habidos en la IA.

Posibles aportaciones de los sistemas expertos a la PC

A pesar de la evidente simplificación que supone la panorámica de los sistemas expertos que hemos intentado trazar, puede ser suficiente para resaltar dos cuestiones. En primer lugar, que los sistemas expertos constituyen un joven campo de la IA que avanza a gran velocidad en lo que a investigación y logros se refiere. En segundo lugar, que los sistemas expertos suponen un modo de resolver problemas mediante computadores cualitativamente distinto del resto de la IA, en el sentido de que se trata de programas con una filosofía y estructura muy peculiares, que entrañan novedades conceptuales importantes. Hasta tal punto esto es así que algunos autores no dudan en afirmar que la aparición y difusión de los sistemas expertos es uno de los cambios más revolucionarios de la historia de la IA (Waterman, 1986; Feigenbaum, 1977, cit. en Hayes-Roth *et al.*, 1983). Parece interesante, por tanto, analizar el posible alcance de este nuevo enfoque desde el punto de vista particular de la PC. Desde este supuesto, vamos a comentar brevemente algunos puntos que, a nuestro juicio, merecen especial atención en tal análisis.

Un problema que aparece constantemente en IA es el «problema de los dominios de juguete» (*toy-domain problem*). Brevemente, consiste en que,

si construimos un programa para resolver un cierto tipo de problemas en un dominio simplificado (como puede ser, por ejemplo, el ya mencionado «mundo de bloques»), los recursos computacionales empleados (representación, mecanismos inferenciales, etc.) dejan de ser cuantitativa y *cualitativamente* válidos para resolver problemas lógicamente análogos, pero a escala real (Waterman, 1986). Podríamos muy bien considerar que el problema de los *toy-domains* es la versión IA del problema psicológico de la validez externa; y equivalente, en psicología cognitiva, a la generalización de los modelos obtenidos de la investigación en microtareas de laboratorio. Es una idea bastante aceptada que los procesos, digamos, atencionales y perceptivos puestos en juego por los sujetos en esas tareas tan específicas no son (al menos, no exactamente) los mismos que emplean en situaciones del mundo, y por ello los modelos resultan difíciles de generalizar. Pero podría tal vez ocurrir que no se tratara sólo de eso, sino que los recursos computacionales requeridos por los dos tipos de situaciones fueran, como ha ocurrido en IA, *cualitativamente* diferentes. De ser así, la generalización de algunos modelos podría ser no ya difícil, sino intrínsecamente imposible, y los procesos postulados serían *esencialmente* artefactos. Una característica de los sistemas expertos es que, por definición, el «problema de los dominios de juguete» no les afecta: se conciben directamente para manejar problemas a escala real. Por tanto, la arquitectura de los sistemas expertos, que bien podría ser menos «elegante» que la de otros sistemas, representa el producto del trabajo de investigadores cuya motivación principal es la eficacia de la actuación del sistema en el mundo real. Y en esta dimensión de «necesidad de eficacia» (Delclaux, 1983; Zaccagnini, 1984; Simon y Zaccagnini, en prensa) constituyen, por tanto, el producto de la IA más próximo a la Psicología [a menos que estemos dispuestos a negar que un sujeto humano tiene, ante todo, que resolver eficazmente problemas reales para (sobre)vivir].

Otra cuestión muy diferente es si ese producto, esa arquitectura computacional, es o no relevante para la comprensión de la inteligencia humana. La arquitectura de un sistema experto se construye en torno a una base de conocimientos, es decir, supone una distinción de raíz entre el «conocimiento» del tema y todos los restantes datos y funciones del sistema. Sería razonable preguntarnos hasta qué punto el contenido de una base de conocimientos se asemeja a lo que la PC entiende por «conocimiento humano». Desgraciadamente, la PC no parece haber producido una definición, psicológica y mínimamente consensuada, de lo que hay que entender por «conocimiento». El término se encuentra en la literatura mezclado e intercambiado con otros como «información», «datos», «representaciones», etc. (Zaccagnini, 1985, 1986, y Sayre, 1986). Se supone que se puede atribuir un contenido de conocimiento a estructuras de información como un modelo mental, una red de la memoria semántica, un esquema, un «proceso» o una imagen mental, pero no existe un criterio claro para decidir cuándo una estructura de información incorpora (o «es») conocimiento. Empírica e intuitivamente, es obvio que poseer toda la información (ya se entienda el término en sentido técnico o vulgar) sobre, por ejemplo, programación de computadores, no es lo mismo que comprender las implicaciones de tal actividad o «saber programar» (y, evidentemente, no se trata aquí de la clásica distinción entre conocimiento declarativo y procedimental). Así pues,

no podemos determinar si el contenido de la base de conocimientos de un sistema experto se parece o no al fenómeno psicológico «conocimiento», por la misma razón que no podemos determinar si la inteligencia artificial es «inteligencia» en el sentido psicológico: en ninguno de ambos casos posee la Psicología una definición lo bastante satisfactoria para servir de criterio a una concepción exógena.

Sí podemos, en cambio, discutir de forma especulativa hasta qué punto esa comparación entre el conocimiento psicológico y el contenido de una base de conocimientos resulta «plausible». Este último consiste en muchos casos en un conjunto parcialmente estructurado de reglas que representan «trucos», índices, estrategias generales y heurísticos que los expertos humanos parecen emplear al resolver los problemas que plantea un campo temático específico. Introspectivamente, no es difícil aceptar la posibilidad de que en muchos casos los sujetos humanos actuemos así. En todo caso, en los sistemas expertos ese contenido de la base de conocimientos contrasta llamativamente con el de la base de hechos, que consiste en un conjunto de información carente de una estructura significativa apreciable. Todo esto nos puede llevar a muchos interrogantes: ¿tendría sentido establecer, en una «hipótesis fuerte», que los trucos, indicios y estrategias implementados en una base de conocimientos *son* conocimiento en un sentido psicológico, es decir, equivalente al que nosotros poseemos y manejamos? ¿Podrían, pues, las estructuras informales de reglas que las constituyen desempeñar un papel semejante al que ya han cumplido las redes semánticas o los «marcos» y «guiones» como constructos psicológicos? ¿Supondrían alguna ventaja respecto a éstos? ¿Podríamos deducir alguna predicción empírica de una hipótesis de esa naturaleza? ¿Podríamos precisar más las nociones psicológicas de «conocimiento» e «información», sobre la base de una distinción semejante a la que se hace entre base de conocimientos y base de hechos?

Desde un punto de vista metodológico, también surgen cuestiones interesantes. Por ejemplo, una de las líneas de trabajo más relevantes de la PC actual se centra en el análisis de las diferencias entre expertos y novatos en tareas que plantean demandas intelectuales. Tales tareas corresponden a líneas de investigación muy variadas (vg. Anderson, *et al.*, 1981; Lewis, 1981 y Pozo, 1987). Podría ser interesante, por ejemplo, aplicar los métodos de la ingeniería del conocimiento a expertos y novatos, y construir bases de conocimientos a partir de ambos. Obtendríamos así representaciones relativamente estructuradas del conocimiento de ambos colectivos de sujetos, las cuales permitirían comparaciones precisas y análisis exhaustivos. Quizá esto sería de más utilidad para la investigación que, por ejemplo, tratar de traducir el conocimiento de los sujetos a formatos representacionales abstractos y generales como el cálculo de predicados, las redes semánticas o los propios sistemas de producciones, los cuales solamente implican restricciones formales, pero no limitaciones derivadas de la necesidad de actuar en un dominio. ¿Podría algún procedimiento de esta naturaleza servir como metodología para el análisis de la adquisición de conocimientos?

Recíprocamente, es posible que la PC pueda aportar mucho a la construcción de sistemas expertos. Ya hemos señalado que los ingenieros del conocimiento se han visto obligados a reconstruir artesanalmente algo muy parecido al método clínico piagetiano, ¿podría éste, tal como se emplea en

la investigación psicológica, ser de utilidad para esos cometidos? Dado que el trabajo del ingeniero del conocimiento implica entrevistas en las que se trata de «extraer» lo que un sujeto humano sabe, probablemente la tecnología desarrollada para fines semejantes por la Psicología (la escuela de Ginebra, la evaluación conductual, etc.) podría ser una ayuda importante.

Si en el área teórico-metodológica cabe hablar de potenciales aportaciones de los sistemas expertos, otro tanto ocurre en la PC aplicada. Por ejemplo, en campos tan actuales como la psicología instruccional o la formación ocupacional, se plantea desde hace tiempo la utilización de los computadores como instrumentos de enseñanza (Educación o Instrucción asistida por ordenador: Coulson, 1962 y Suppes, 1975). Tanto las ventajas como las limitaciones de este enfoque son bien conocidas, y las hemos discutido en otros lugares (Adarraga, 1985; Zaccagnini *et al.*, 1985). Uno de los problemas generales más serios que se presentan es la relativa rigidez del *software* para una situación instruccional en la que el sujeto que interactúa con el ordenador resulta muy difícilmente previsible en su comportamiento, su facilidad o dificultad para el aprendizaje de unos u otros conceptos, etc. Existen ya hoy numerosos proyectos de IA educativa, alguno de los cuales se sirven de programas con la arquitectura de los sistemas expertos. En principio, parece que buena parte de esa rigidez de los programas puede eliminarse utilizando un sistema que tiene «conocimientos» del tema a enseñar, y que puede presentarlos acomodándose a situaciones imprevistas: ya a partir de MYCIN se diseñó GUIDON, sistema que tiene la función de entrenar a estudiantes de medicina en las habilidades diagnósticas de aquél. En este sentido, la arquitectura computacional de los sistemas expertos permite estructuras muy prometedoras: ya se han diseñado partiendo de ellas sistemas que incorporan, además de una base de conocimientos sobre el tema a enseñar, un contenido explícito de conocimientos pedagógicos y la posibilidad de construirse, a lo largo del proceso de instrucción, un «modelo» dinámico del alumno (véase, por ejemplo, Sleeman y Brown, 1982).

Por último, cabe preguntarse (o más bien, «meta-preguntarse») hasta qué punto estas cuestiones representan posibilidades reales para la PC, o son mera «ciencia-ficción» especulativa. Desde luego, no creemos en absoluto que los sistemas expertos sean una suerte de «panacea» para los problemas de la PC; sin embargo, parece evidente que algunas de las vías que ofrecen son realmente relevantes para algunos de éstos, y merecen una consideración más detallada. Por ello, esperamos hacerlas objeto de un próximo trabajo.

Referencias

- ADARRAGA, P. (1985). Criterios educacionales en la elección de software. En: Pfeiffer, A. y Galvan, J. (comps.): *Informática y Escuela*. Madrid: Fundesco.
- ADARRAGA, P. (1986). Eficacia de diferentes estrategias en el procesamiento mnésico de información verbal: la hipótesis de la funcionalidad específica. Memoria de Licenciatura. Madrid: UAM.
- ALEKSANDER, I. (1986). Artificial intelligence and training: an initial assessment. En: CEDEFOP (comp.): *Artificial Intelligence and its applications to vocational training*. Berlín: CEDEFOR.

- ANDERSON, J. R. (1976). *Lenguaje, memory and thought*. Hillsdale: Lawrence Earlbaum Assoc.
- ANDERSON, J. R. (1983). *The architecture of cognition*. Harvard University Press.
- ANDERSON, J. R.; GREENO, J. G.; KLINE, P. J., y NEVES, D. M. (1981). Acquisition of problem solving skill. En: J. R. Anderson (comp.): *Cognitive skills and their acquisition*. Hillsdale: Lawrence Earlbaum Assoc.
- BARR, A. y FEIGENBAUM, E. A. (1982a). *The Handbook of Artificial Intelligence. Vol. I*. Los Altos: William Kaufmann.
- BARR, A. y FEIGENBAUM, E. A. (1982b). *The Handbook of Artificial Intelligence. Vol. II*. Los Altos: William Kaufmann.
- BARSTOW, D. R.; AIELLO, N.; DUDA, R. O.; ERMAN, L. D., et al. (1983). Languages and tools for knowledge engineering. En: F. Hayes-Roth, D. A. Waterman y D. B. Lenat (comps.): *Building expert systems*. Addison-Wesley.
- BODEN, M. (1977). *Artificial intelligence and natural man*. Harvester Press.
- BODEN, M. (1981). *Minds and mechanisms*. The Harvester Press.
- BELINCHÓN, M. (1987). Lenguaje. En J. M. Ruiz-Vargas (comp.): *Esquizofrenia: un enfoque cognitivo*. Madrid: Alianza Psicología.
- BUCHANAN, B. G. y FEIGENBAUM, E. A. (1982). Prólogo a Davis y Lenat. 1982.
- BUCHANAN, B. G. y SHORTLIFFE, E. H. (comps.) (1982). *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley.
- CARELLO, C.; TURVEY, M. T.; KUGLER, P. N., y SHAW, R. E. (1984). Inadequacies of the computer metaphor. En: M. S. Gazzaniga (comp.): *Handbook of cognitive neuroscience*. Nueva York: Plenum.
- CARRETERO, M. y GARCÍA MADRUGA, J. A. (comps.) (1982). *Lecturas de psicología del pensamiento*. Madrid: Alianza Psicología.
- CHAMBERLAIN, W. (1984). *The policeman's beard is half-constructed*. Warner Books.
- CHANG, C. y LEE, R. C. (1973). *Symbolic logic and mechanical theorem proving*. Nueva York: Academic Press.
- COHEN, P. P. y FEIGENBAUM, E. A. (1982). *The Handbook of Artificial Intelligence. Vol. III*. Los Altos: William Kaufmann.
- COULSON, R. (comp.) (1962). *Programmed learning and computer-based instruction*. J. Wiley & Sons.
- CUENA, J. (1984). Sistemas expertos. En: R. Valley, J. Barbera y F. Ros (comps.). *Inteligencia artificial: introducción y situación en España*. Madrid: Fundesco.
- CUENA, J. (1986a). Introducción general a la Inteligencia Artificial. En: J. Cuenca (comp.): *Inteligencia artificial: Sistemas expertos*. Madrid: Alianza Informática.
- CUENA, J. (1986b). Adquisición de conocimiento y aprendizaje en sistemas basados en reglas. En J. Cuenca (comp.): *Inteligencia artificial: Sistemas expertos*. Madrid: Alianza Informática.
- DAVIS, R. y KING, J. (1977). An overview of production systems. En E. W. Elcock y D. Michie (comps.): *Machine intelligence 8*. J. Wiley & sons.
- DAVIS, R. y LENAT, D. B. (1982). *Knowledge-based systems in artificial intelligence*. McGraw-Hill.
- DEHN, N. y SCHANK, R. C. (1982). Artificial and human intelligence. En: R. J. Sternberg (comp.): *Handbook of Human Intelligence*. Cambridge University Press.
- DENNET, D. C. (1981). Intentional systems. En: J. Haugeland (comp.): *Mind design. Philosophy, Psychology, Artificial Intelligence*.
- DENNET, D. C. (1983). *Intentional systems in cognitive ethology: the «panglossian paradigm» defended*. *The Behavioral and Brain Sciences* 6 (3), 343-390.
- DELCLAUX, I. (1983). La eficacia en la actuación humana. En: I. Delclaux (comp.): *La intervención psicológica: Sistemas de producción y consumo*. Murcia: Ed. Límites.
- DELCLAUX, I. (1982). Introducción al procesamiento de información en Psicología. En: I. Delclaux y J. Seoane (comps.): *Psicología cognitiva y procesamiento de la información*. Madrid: Pirámide.
- DELCLAUX, I. y SEOANE, J. (comps.) (1982). *Psicología Cognitiva y Procesamiento de la Información*. Madrid: Pirámide.
- DRETSKE, F. I. (1981). *Knowledge and the flow of information*. Bradford Books.
- FERNÁNDEZ, G. (1986). Panorama de los sistemas expertos. En: J. Cuenca (comp.): *Inteligencia artificial: Sistemas expertos*. Madrid: Alianza Informática.
- FODOR, J. A. (1966). Could there be a theory of perception?. Trad. en: García-Albea (1986b). *Percepción y computación*. Madrid: Pirámide.
- FODOR, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioral and brain sciences*, 3 (1), 63-109.
- FODOR, J. A. (1983). *The modularity of mind*. MIT Press. Trad.: *La modularidad de la mente*. Madrid: Morata, 1986.
- FODOR, J. A. (1985). Précis of The Modularity of Mind. *The Behavioral and Brain Sciences* 8, 1-42.
- GAMBARA, H. y LEÓN, O. (1988). Uso de la información en el proceso de toma de decisiones. *Boletín de Psicología*, 17 (en prensa).

- GARCÍA-ALBEA, J. E. (1986a). Prólogo a la edición en castellano de J. A. Fodor: *La modularidad de la mente*. Madrid: Morata.
- GARCÍA-ALBEA, J. E. (comp.) (1986a). *Percepción y computación*. Madrid: Pirámide.
- GARDNER, M. (1982). *Logic Machines and Diagrams*. Chicago: University of Chicago Press. Trd.: *Máquinas y diagramas lógicos*. Madrid: Alianza.
- GARDNER, M. (1985). *The mind's new science. A history of the cognitive revolution*. Nueva York: Basic Books.
- GIBSON, J. J. (1979). *The ecological approach to visual perception*. Nueva York: Houghton Mifflin.
- HAUGELAND, J. (1981a). Semantic engines. En: J. Haugeland (comp.): *Mind design. Philosophy, Psychology, Artificial Intelligence*. Bradford.
- HAUGELAND, J. (1981b). The nature and plausibility of cognitivism. En: J. Haugeland (comp.): *Mind design. Philosophy, Psychology, Artificial Intelligence*. Bradford.
- HAYES-ROTH, F.; WATERMAN, D. A., y LENAT, D. B. (comps.) (1983). *Building Expert Systems*. Addison-Wesley.
- 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.
- KOWALSKI, R. y KUCHNER, D. (1971). Linear resolution with selector function. *Artificial Intelligence*, 2, 227-260.
- KYBURG, H. E. (1983). Rational belief. *The Behavioral and Brain Sciences*, 6 (2), 231-273.
- LENAT, D. B. (1982). The nature of heuristics. *Artificial Intelligence*, 19, 189-249.
- LENAT, D. B. (1982). EURISKO: A program that learns new heuristics and domain concepts. The nature of heuristics III: program design and results. *Artificial Intelligence*, 21, 61-98.
- LENAT, D. B. (1984). Programación de sistemas inteligentes. *Investigación y Ciencia*, 98, 140-149.
- LENAT, D. B. y BROWN, J. (1984). Why AM and EURISKO appear to work. *Artificial Intelligence*, 23, 269-294.
- LEWIS, C. (1981). Skill in algebra. En: J. R. Anderson (comp.): *Cognitive skills and their acquisition*. Lawrence Erlbaum Ass.
- LOVELAND, D. (1978). *Automatic theorem-proving: a logical basis*. Amsterdam: North-Holland.
- MCCORDUCK, P. (1979). *Machines who think*. San Francisco: W. H. Freeman.
- MCDERMOTT, D. (1976). Artificial intelligence meets natural stupidity. *SIGART Newsletter (ACM)* 57. Reimp. en Haugeland, 1981.
- MARR, D. (1978). Representing visual information. En: A. R. Hanson y E. M. Riseman (comps.): *Computer vision systems*. Nueva York: Academic Press.
- MARR, D. (1986a). *Procesamiento de la información visual: estructura y creación de las representaciones visuales*. En: J. E. García-Albea (1986b): *Percepción y computación*. Madrid: Pirámide.
- MARR, D. (1986b). *La visión*. Madrid: Alianza Psicología.
- MICHALSKI, R. S.; CARBONELL, J. G., y MITCHELL, T. M. (1986). *Machine Learning: An artificial intelligence approach*. Palo Alto: Tioga Publ.
- MINSKY, M. (1975). A framework for representing knowledge. Rep. en: J. Haugeland (comp.) (1981): *Mind Design*. Bradford Books.
- MITCHELL, T. M.; CARBONELL, J. G., y MICHALSKI, R. S. (comps.) (1983). *Machine Learning: A guide to current research*. Boston: Kluwer Academic Publ.
- NAUGHTON, J. (1986). Artificial intelligent and its applications to vocational training. Extracto de un informe de la Open University. En: CEDEFOP (comp.): *Artificial Intelligence and its applications to vocational training*. Berlín: CEDEFOP.
- NAYLOR, C. (1983). *Build your own expert system*. Cheshire: Sigma Technical Press. Trad.: *Construya su propio sistema experto*. Madrid: Díaz de Santos, 1986.
- NEGOITA, C. V. (1985). *Expert systems and fuzzy systems*. Addison-Wesley.
- NEISSER, U. (1976). *Cognition and reality*. Trad.: *Procesos cognitivos y realidad*. Madrid: Marova.
- NEWELL, A. (1962). Some problems of basic organization in problem-solving programs. En: A. Yovits, G. T. Jacobi y G. D. Goldstein (comps.): *Self-organizing systems*. Nueva York: Spartan.
- NEWELL, A. (1970). Remarks on the relationship between Artificial Intelligence and Cognitive Psychology. En: R. Banerji y M. D. Mesarovic (comps.): *Theoretical approaches to non numerical problem-solving*. Springer Verlag.
- NEWELL, A. (1973a). Artificial intelligence and the concept of mind. En R. C. Schank y K. Colby (comps.): *Computer models of thought and language*. W. H. Freeman.
- NEWELL, A. (1973b). Production systems: Models of control structures. En: W. Chase (comp.): *Visual information processing*. Nueva York: Academic Press.
- NEWELL, A. (1980). Physical symbol systems. *Cognitive Science* 4 (2), 135-183.
- NEWELL, A. y SIMON, H. A. (1972). *Human problem solving*. Prentice-Hall.
- NEWELL, A. y SIMON, H. A. (1981). Computer science as empirical inquiry: symbols and search. En: J. Haugeland (comp.): *Mind desing. Philosophy, Psychology, Artificial Intelligence*. Bradford Books.

- NILSSON, N. J. (1982). *Principles of Artificial Intelligence*. Springer-Verlag.
- PEARL, J. (1984). *Heuristics: Intelligent search strategies for computer problem solving*. Addison-Wesley.
- PIAGET, J. (1981). *La representación del mundo en el niño*. Madrid: Morata.
- POHL, I. (1977). Practical and theoretical considerations in heuristic search algorithms. En E. W. Elcock y D. Michie (comps.): *Machine Intelligence, 8*. Nueva York: Wiley.
- POST, E. (1943). Formal reduction of the general combinatorial problem. *American Journal of Mathematics, 65*, 197-268.
- POZO, J. I. (1987). *Aprendizaje de la ciencia y pensamiento causal*. Madrid: Visor.
- PUTMAN, H. (1973). Reductionism and the nature of psychology. *Cognition, 2*, 131-146.
- PYLYSHYN, Z. W. (1980). Cognition and computation: Issues in the foundations of Cognitive Science. *The Behavioral and Brain Sciences, 3* (1), 111-132.
- PYLYSHYN, Z. W. (1984). *Computation and cognition. Toward a foundation for cognitive science*. Bradford.
- QUILLIAN, M. R. (1968). Semantic memory. En: M. Minsky (comp.): *Semantic information processing*. MIT Press.
- QUILLIAN, M. R. (1979). Discovering rules from large collections of examples. A case study. En: D. Michie (comp.): *Expert systems in the microelectronic age*. Edinburgh Univ. Press.
- RIEDL, R. (1983). *Biología del conocimiento. Los fundamentos filogenéticos de la razón*. Barcelona: Labor.
- RIVIÈRE, A. (1984). La psicología de Vygotski: sobre la larga proyección de una corta biografía. *Infancia y Aprendizaje, 27/28*, 7-86.
- RIVIÈRE, A. (1986). *Razonamiento y representación*. Madrid: Siglo XXI.
- RIVIÈRE, A. (1987). *El sujeto de la Psicología Cognitiva*. Madrid: Alianza Psicología.
- ROSCH, E. (1978). Principles of categorization. En: E. Rosch y B. B. Lloyd (comps.): *Cognition and categorization*. Lawrence Earlbaum Ass.
- ROSCH, E. y MERVIS, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology, 7*, 575-605.
- ROSENBLOOM, P. S.; LAIRD, J. G.; NEWELL, A., et al. (1986). Current research on learning in SOAR. En: M. Mitchell et al. (comps.) (1986): *Machine Learning: An artificial intelligence approach*. Palo alto: Tioga Publ.
- ROSENFELD, A. y KAK, A. C. (1976). *Digital picture processing*. Nueva York: Academic Press.
- RUIZ-VARGAS, J. M. y ZACCAGNINI, J. L. (1987). El modelo del Procesamiento de información. En: J. M. Ruiz-Vargas (comp.): *Esquizofrenia: un enfoque cognitivo*. Madrid: Alianza Psicología.
- SAYRE, K. M. (1986). *Intentionality and information processing: an alternative model for cognitive science*. *The Behavioral and Brain Sciences, 9*, 121-166.
- SCHANK, R. C. (1973). Identification of the conceptualizations underlying natural language. En: R. C. Schank y K. Colby (comps.): *Computer models of thought and language*. W. H. Freeman.
- SCHANK, R. C. y COLBY, (comps.) (1973). *Computer models of thought and language*. W. H. Freeman.
- SCHANK, R. C. y ABELSON, R. P. (1977). *Scripts, plans, goals and understanding*. Hillsdale, Nueva York: Lawrence Earlbaum Assoc.
- SCHNEIDERMAN, B. (1986). *Designing the user interface. Strategies for effective human-computer interaction*. Addison-Wesley.
- SEOANE, J. (1979). Inteligencia artificial y procesamiento de información. *Boletín informativo de la Fundación Juan March, 85*, 3-21.
- SHORTLIFFE, E. H. (1976). *Computer-based medical consultations: Mycin*. Nueva York: American Elsevier.
- SIMON, C. y ZACCAGNINI, J. L. (en preparación). Evaluación de las estrategias de adquisición de conocimientos en estudiantes universitarios.
- SIMON, H. A.; LANGLEY, P. W., y BRADSHAW, G. L. (1981). Scientific discovery as problem solving. *Synthese, 47*, 1-27.
- SLEEMAN, D. y BROWN, J. S. (comps.). *Intelligent tutoring systems*. London: Academic Press.
- STEFIK, M.; AIKINS, J.; BALZER, R.; BENOIT, J., et al. (1983a). Basic concepts for building expert systems. En: F. Hayes-Roth, D. A. Waterman y D. B. Lenat (comps.) (1983): *Building expert systems*. Addison-Wesley.
- STEFIK, M.; AIKINS, J.; BALZER, R.; BENOIT, J., et al. (1983b). The architecture of expert systems. En: F. Hayes-Roth, D. A. Waterman y D. B. Lenat (comps.) (1983): *Building expert systems*. Addison-Wesley.
- SUPPES, P. (1975). Los usos de las computadoras en la educación. En: R. C. Atkinson (comp.): *Psicología contemporánea*. Madrid: H. Blume.
- TVERSKY, A. y KAHNEMAN, D. (1982). Juicios en situación de incertidumbre: heurísticos y sesgos. En: M. Carretero y J. A. García-Madruga (comps.): *Lecturas de psicología del pensamiento*. Madrid: Alianza Psicológica.
- VALIANT, L. G. (1986). What can be learned?. En: T. M. Mitchell, J. G. Carbonell y R. S. Michalski (comps.): *Machine learning: A guide current research*. Boston: Kluwer Academic Publ.

- DE VEGA, M. (1981). Una exploración de los metapostulados de la psicología contemporánea: El logicismo. *Análisis y Modificación de la Conducta*, 7, 345-375.
- DE VEGA, M. (1982). La metáfora del ordenador: Implicaciones y límites. En: I. Delclaux y J. Seoane (comps.): *Psicología cognitiva y procesamiento de la información*. Madrid: Pirámide.
- DE VEGA, M. (1983). Filogénesis, adaptación y sesgos biológicos del conocimiento: Una alternativa de las analogías formales. *Boletín de Psicología*. Cátedra de Psicología Social. Universidad de Valencia.
- DE VEGA, M. (1987). Re-procesamiento de la información: Una breve incursión en los abismos informacionales. *Boletín de Psicología*, 15, 7-12.
- VERDEJO, M. F. (1986). Sistemas basados en reglas de producción y programación lógica. En: J. Cuenca (comp.): *Inteligencia artificial: Sistemas expertos*. Madrid: Alianza Informática.
- WALTZ, D. (1982). Artificial intelligence. *Scientific American*, 247, 118-133.
- WATERMAN, D. A. (1986). *A guide to expert systems*. Addison-Wesley.
- WATERMAN, D. A. y HAYES-ROTH, F. (1983). An investigación of tools for building expert systems. En: F. Hayes-Roth, D. A. Waterman y D. B. Lenat (comps.): *Building expert systems*. Addison-Wesley.
- WEIZENBAUM, J. (1976). *Computer power and human reason. From judgement to calculation*. W. H. Freeman.
- WINOGRAD, T. (1972). *Understanding natural language*. Nueva York: Academic Press.
- WINOGRAD, T. (1973). *Language as a cognitive process. Vol. I: Syntax*. Addison-Wesley.
- ZACCAGNINI, J. L. (1984). *Estrategias mnésicas en el procesamiento de información verbal*. Tesis Doctoral. Madrid: UAM.
- ZACCAGNINI, J. L. (1985). «Modelos de Mundo» como marco de referencia para una psicología del conocimiento. *Estudios de Psicología*, 22, 19-32.
- ZACCAGNINI, J. L. (1986). *Proyecto docente para Psicología Básica (III): Memoria*. Madrid: UAM. Memoria no publicada.
- ZACCAGNINI, J. L. y DELCLAUX, I. (1982). Psicología Cognitiva y procesamiento de la información. En: I. Delclaux y J. Seoane (comps.): *Psicología cognitiva y procesamiento de la información*. Madrid: Pirámide.
- ZACCAGNINI, J. L. y MORALES, F. (1985). La «Nueva Psicología Social Cognitiva» desde una perspectiva «Cognitiva». *Revista de Psicología General y Aplicada*, 40 (4), 727-755.
- ZACCAGNINI, J. L.; VILLEGAS, P.; SANZ, S.; PANIAGUA, A., y ADARRAGA, P. (1985). *Estudio sobre normas y criterios para la construcción y evaluación de software*. Fundesco.