

LA AUTOMATIZACIÓN DEL RAZONAMIENTO ARQUEOLÓGICO

Juan A. Barceló¹

INTELIGENCIA ARTIFICIAL Y PRÁCTICA ARQUEOLÓGICA

Aunque la utilización de las técnicas de Inteligencia Artificial no está muy difundida en la Arqueología actual, podemos mencionar algunos ámbitos de nuestra disciplina que han de beneficiarse, necesariamente, con los actuales avances en tecnologías de programación de ordenadores:

Los análisis estadísticos (uni- y multidimensionales) no tardarán en ser sustituidos por las técnicas de Inducción Automática, utilizadas en la actualidad como útil exploratorio en grandes bases de datos. Ventajas: facilidad de uso, sencillez para interpretar el resultado del análisis (Reglas vs. Ecuaciones), ausencia de parametrización y no dependencia con la Curva Normal (cf. Barceló 1992a, Weiss y Kulikowski 1991).

Tipologías Automáticas: El uso de Sistemas Expertos constituye la alternativa más evidente a las listas tipológicas, ya que permiten acumular más datos, su lectura no es ambigua, y gestionan incluso las excepciones (Benfer et al., 1991, Markel 1987, Maicas, 1989).

Tratamiento de imágenes. En Arqueología todo es imagen, desde la microfotografía del desgrasante de una cerámica hasta la planimetría de una región, pasando por la forma y decoración de los artefactos, la disposición micro-espacial del registro espacial y las secciones estratigráficas. Hasta ahora esas imágenes tenían que traducirse en un lenguaje descriptivo que en la mayoría de los casos demostraba ser limitado por su dependencia con respecto a la subjetividad del que hacía la descripción. El uso de Redes Neuronales y Sistemas Distribuidos permitirá, a partir de ahora, el análisis directo de las imágenes, sin tener que pasar por la fase de la codificación descriptiva en variables y atributos (Zeidenberg 1990, Gibson 1993, Claxton 1993).

No obstante estos avances “reales” en el ámbito *práctico*, algunos proyectos recientes intentan ir más allá, pretendiendo *simular* la manera en que piensan los arqueólogos (cf. por ejemplo: Doran 1990, Reynolds 1986, Francfort et al., 1989, Stutt y Shennan 1991, Barceló 1992b, 1993). El propósito de

1. *Universitat Autònoma de Barcelona*, Departamento de Antropología Social y Prehistoria.

estos programas es estudiar los mecanismos cognitivos subyacentes en el razonamiento arqueológico, *visualizando* el razonamiento en sí.

Todos estos proyectos se inscriben de lleno en lo que podríamos denominar *Arqueología Automática*, : una subdisciplina encargada de poner de manifiesto la manera en que los arqueólogos interpretan sus datos. El razonamiento científico es un mecanismo inobservable, del cual sólo conocemos sus efectos (las interpretaciones). Por consiguiente, un método de *visualizar* el mecanismo de producción de las interpretaciones debiera ser bienvenido. La Arqueología Automática utiliza medios computacionales para cumplir ese objetivo. Su supuesto básico es el siguiente:

si somos capaces de programar un ordenador como si fuese una persona, empezaremos a comprender cómo actúa esa persona.

Algunos pueden pensar, sin embargo, que un programa “simulador” actúa de manera sensiblemente distinta a la del arqueólogo. La diferencia, sin embargo, ha de entenderse en los términos de una Teoría de la Representación: el programa informático hace las veces de *modelo* o *representación* del mecanismo de producción de interpretaciones arqueológicas. En este artículo intentaré desarrollar algunas de las consecuencias que se derivan de la obtención y del uso de esos modelos de razonamiento.

LA NATURALEZA DE LOS PROBLEMAS ARQUEOLÓGICOS

En los últimos años, matemáticos e informáticos se han dedicado a la formalización de la noción de PROBLEMA (cf. Newell y Simon 1972, Sacerdoti 1977, Laurière 1986, Gilhooly1989). De estos trabajos se derivan importantes consideraciones a tener en cuenta a la hora de plantearnos los correspondientes problemas arqueológicos.

Empezaremos con una definición sencilla: “un problema es una dificultad que no puede resolverse automáticamente, sino que requiere una investigación conceptual o empírica”. En otras palabras, nos planteamos un problema en cuanto nos hallamos en una situación en la que queremos obtener o hacer algo y no conocemos las acciones que hay que emprender para obtener lo que queremos o hacer lo que deseamos.

Si aplicásemos esta idea general al ámbito científico, tendríamos que concluir que *toda interpretación, todo significado, no son más que la solución a un determinado problema*. Sólo hay un modo de resolver este tipo de problemas: disponiendo de un conjunto de interpretaciones posibles alternativas (p.e. cronologías que puede tener un objeto arqueológico) y decidiendo cuál de ellas es la más apropiada en el caso en cuestión, según ciertos criterios bien especificados. En otras palabras,

dado un dominio D, es necesario encontrar en un conjunto X de soluciones posibles, los elementos x que cumplan un conjunto de condiciones K(x) bien especificadas y definidas por el propio enunciado del problema.

D hace referencia al tema o el dominio del saber en el cual se ha planteado el problema, e incluye todos los hechos, axiomas, e hipótesis conocidos; en otras palabras, se trata de una *teoría científica*. Este conjunto suele denominarse *generador* o *espacio* del problema.

El *estado inicial* de un problema coincide con lo que los lógicos denominan *explanans*, esto es, aquello que queremos interpretar. El estado inicial de un problema arqueológico es obvio: el registro

arqueológico, el artefacto o la disposición de artefactos que queremos interpretar, y lo representaremos por medio de una lista de rasgos descriptivos.

La *interpretación* arqueológica coincide con la *solución* al problema arqueológico que hemos planteado. En este caso, hacemos referencia al *explanandum*, que suele denominarse *estado final* del problema. Al igual que el estado inicial, ese estado final no consiste en una simple proposición lingüística del tipo “siglo XX”, sino en el conjunto de características de ciertos objetos del siglo XX: “blancos, con forma de tubo, malolientes, rellenos de hojas de la planta Nicotina Tabacum finamente picadas”, esto es, aquellas características que diferencian esa solución particular de las demás posibles. Cuanto más generales sean los términos que describen esa solución, tanto más amplio será su ámbito de aplicación, y responderá a más problemas.

Ahora bien, para que una unidad de conocimiento así caracterizada sea realmente una solución precisamos de un conjunto de criterios de validación o requisitos para aceptarla como válida; ese conjunto de requisitos constituye el *objetivo*, que puede ser definido en términos sencillos como: “aquello que desea obtenerse”. Por su parte, los *operadores*, suelen definirse como “unidad de conocimiento necesaria para alcanzar un fin”. De ahí que todo objetivo esté representado por la detección de una “falta de información” para obtener algo, en tanto que los operadores estarían representados por el “uso” de cierta información para poder llenar ese vacío de conocimiento.

Aceptando la identidad del mecanismo de producción de interpretaciones con el mecanismo de resolución de problemas, afirmaremos que los datos arqueológicos se interpretan por medio de una operación de *búsqueda*: tanto el arqueólogo como el programa de ordenador que resuelve automáticamente problemas, son capaces de *buscar* la interpretación o estado final que corresponde a la descripción empírica o estado inicial facilitado por el usuario. Según el tipo de problema, la operación de búsqueda será distinta:

- *Problemas de Diagnóstico* (cf. Torasso y Console 1989): se parte de un conjunto de interpretaciones alternativas totalmente conocido y se trata de seleccionar una de ellas, utilizandopara ello diversos criterios que reduzcan la búsqueda a unos límites computables.

- *Problemas de Diseño* (cf. Brown y Chandrasekaran 1989): se trata de construir la interpretación de ciertos datos a partir de una serie de restricciones que deben ser satisfechas. El conjunto de interpretaciones posibles es, en este caso, desconocido, y hay que generarlo basandose en las restricciones. La operación de búsqueda, por tanto, no se realiza entre las interpretaciones candidatas, sino entre los operadores –cuál es el más conveniente para transformar el estado actual en un estado intermedio que cumpla los requisitos–.

Así pues, computacionalmente hablando, la interpretación de un artefacto arqueológico empezará siempre con la identificación de la dificultad que debe resolverse (el *objetivo*); por ejemplo:

¿Cuál es la Cronología de (x)?

¿Qué función desempeñaban los artefactos (y),(z)?

¿Qué estructura social corresponde a la necrópolis (v)?

Una vez definido el objetivo recurrimos a la información disponible (conocimiento previo) y construimos el estado inicial sobre el cual aplicaremos el operador o los operadores necesarios. Recuerdese que la *descripción* no es una operación neutra, sino que depende, no sólo de la teoría a nuestra disposición, sino de los objetivos que hayamos formulado.

A continuación estimamos la diferencia existente entre nuestro punto de partida y el punto al que queremos llegar (el objetivo). Dado que ambos son unidades de conocimiento representadas de la misma manera, esa estimación, aunque compleja debido a las distintas modalidades, es computacionalmente factible. El procedimiento u operador que seleccionemos para *generar* la interpretación tendrá que reducir esa diferencia, permitiéndonos ir del punto de partida hasta el punto de llegada o interpretación.

No hay manera de construir un operador que busque en un conjunto infinito de soluciones o interpretaciones posibles, porque esa búsqueda no acabaría nunca. En Inteligencia Artificial denominamos a esa dificultad *explosión combinatoria*, y suele resolverse recurriendo a una serie de condiciones o *heurísticas* que progresivamente limitarán la cantidad de operadores posibles y que, finalmente, permitirán elegir el más adecuado (Newell y Simon 1972, Pearl 1985). El procedimiento es aparentemente simple: en cada una de las etapas de la resolución se compara el estado actual del problema con el estado final del mismo (si es conocido) o con ciertas condiciones previamente enunciadas que debe cumplir ese estado para poder ser considerado la solución; a continuación se lleva a cabo una operación que permita reducir la diferencia entre el estado actual y las características del estado final. El uso de este procedimiento heurístico permite al agente encargado de resolver el problema avanzar reduciendo progresivamente la distancia a la solución y evitando el uso de estados del problema que aumenten esa diferencia.

Una búsqueda heurística es un procedimiento que aplica pasos plausibles para navegar en un espacio de alternativas inmenso, con el fin de conocer la mejor (o una colección de las mejores) alternativas para cierto propósito. Lo que convierte a esta búsqueda en *heurística* es que el procedimiento no garantiza que el resultado sea la mejor de todas las alternativas, o una colección que incluya la mejor alternativa, si bien el procedimiento se aproximará bastante a ella, de acuerdo con cierto criterio. Ese criterio puede que no sea muy riguroso y que no siempre proporcione una solución óptima, si bien ésta será la mejor que se puede obtener, muy superior a la que se habría logrado sin una búsqueda heurística. Se insiste en aquello que es alcanzable y suficientemente bueno, prescindiendo de lo que es óptimo, pero inalcanzable.

La idea de *heurística* aplicada al razonamiento científico no tiene nada de insólito: ciertas HIPOTESIS VEROSIMILES organizan el proceso de resolución de un problema científico, reduciendo el espacio de dicho problema y la búsqueda en él de un estado final, si éste es desconocido. Si el estado inicial fuese, por ejemplo, la descripción de un yacimiento arqueológico y de los objetos que en él han aparecido, los estados posibles del problema arqueológico son infinitos, con lo que nunca se llegará a una solución. Mas si proponemos una HIPOTESIS, esto es, una Función de la Cultura Material en ese yacimiento y su Contexto de uso habríamos limitado drásticamente el conjunto de interpretaciones posibles. La tarea sería entonces *construir* el mecanismo de conexión entre el estado inicial y el estado final que proponemos. La diferencia con la visión positivista de la demostración de hipótesis radica, precisamente, en la “posibilidad”, antes que en la “probabilidad” de esa construcción: no descubrimos una conexión real, sino que proponemos una demostración plausible. En el momento en que cambiamos las condiciones que hemos impuesto en el espacio del problema, el resultado será otro, impredecible.

La racionalidad científica consiste, en usar los mejores medios que permiten reducir la búsqueda de una solución *acceptable* a unas proporciones manejables. Esos medios heurísticos sugieren qué operadores hay que aplicar en cada momento, cuál es el estado del problema que más nos acerca

a la solución y cómo *generar* ese estado –esto es, esa unidad de información particular– si no la conocemos previamente, cómo han de evaluarse los estados sucesivos de un problema,... Es decir, la forma de utilizar la información necesaria para resolver un problema está controlada, de un modo u otro, por lo que sabemos acerca de ese problema. Esa es la única forma de poder llegar a una solución en un plazo de tiempo razonable: el conocimiento del problema nos permite estipular las condiciones que han de cumplir los estados sucesivos del problema para llegar a ser una solución aceptable al mismo. Si la información disponible es muy escasa, entonces se usarán ciertos conocimientos generales acerca de la solución más aconsejable para proponer nuevas hipótesis (forzosamente muy generales en un primer momento) y contrastarlas (parcialmente).

De todo lo dicho hasta aquí se deduce que la solución a un problema –ya sea arqueológico o propio de la vida cotidiana– surge de la manipulación de un conjunto de unidades de conocimiento. Cuanto más numeroso y mejor estructurado esté ese conjunto de unidades de conocimiento, más se parezca a una Teoría Científica y más integrado esté el operador en esa Teoría, más fácil será resolver el problema y más adecuadas las interpretaciones que obtengamos. Por consiguiente, definiremos el término *significado arqueológico* como “el conjunto de acciones sucesivas que conducen de la descripción empírica de un fenómeno arqueológico a su *interpretación*. Esas “acciones” son, en realidad, la aplicación de distintos operadores cuya función es poner en contacto el *estado inicial* (descripción del fenómeno) con el *estado final* (o interpretación considerada válida).

LA NATURALEZA DEL RAZONAMIENTO ARQUEOLÓGICO

Ya hemos presentado el propósito fundamental de la Arqueología Automática: *visualizar la mecánica del razonamiento arqueológico por medio de útiles de programación*. Esa “visualización” es sencilla partiendo del mecanismo básico de resolución automática de problemas, el cual se reduce, en realidad, a un conjunto de asociaciones entre conceptos. De este modo, las inferencias arqueológicas están contenidas en los enlaces y conexiones entre estados intermedios que conducen del estado inicial al estado final de un problema; esas asociaciones han sido producidas, evidentemente, gracias a la aplicación de uno o varios operadores. Este formato de representación debiera permitirnos, además, definir adecuadamente la noción misma de *razonamiento*.

En Inteligencia Artificial, el espacio del problema, esto es, el conjunto de todas las interpretaciones posibles (la *teoría*) suele representarse por medio de un grafo (Sowa 1984,1991, Shastri 1988, Frixione et al., 1992). Los distintos estados del problema adoptan la forma de los nodos de dicho grafo, en tanto que los arcos deben entenderse en tanto que *asociaciones* entre dos estados consecutivos, y constituyen, por tanto, el resultado de la acción de cierto operador. Denominaremos *cadena de inferencias* a una secuencia de nodos y arcos en un grafo que representen la relación de dependencia que existe entre la unidad que representa la descripción de cierto artefacto o conjunto de artefactos (estado inicial) y la unidad que representa su interpretación (estado final). Es importante tener presente que las distintas unidades incluídas en la cadena de inferencias son muy diversas entre sí –la unidad que representa la interpretación, por ejemplo, suele ser más *general* y *abstracta* que la unidad que hace las veces de observación–, de ahí que el tipo de asociaciones que establezcamos entre ellas tendrán que dar cuenta de esas diferencias, especialmente en lo que se refiere al orden jerárquico entre las mismas.

Así pues, “razonar” equivale a asociar o combinar de determinada manera una serie de unidades de conocimiento, también llamadas *conceptos*. Cualquier “razonamiento”, esto es, cualquier resultado de

la acción de razonar, puede expresarse por medio de cadenas de inferencia. Existen dos formas de representar una *cadena de inferencias* (o “esquema de asociaciones”):

- *proposicionalmente*, con ayuda de expresiones pseudo-lingüísticas

- *topológicamente*, mediante el análisis de los invariantes en las correspondencias puntuales biunívocas y continuas que definen el grafo del problema.

En el primer caso utilizaremos como vehículo de representación las denominadas *Reglas de Producción*, que expresan la asociación entre los estados del problema bajo la forma de pares Estimulo-Respuesta. En otras palabras, lo que se propone es el uso de un conjunto finito de reglas sencillas, como

si A entonces haz B

si A' entonces x es B'

En ellas, la parte izquierda (SI A...) es la condición de aplicación, y la parte derecha (ENTONCES...B) la acción o *generador* de la interpretación. Superficialmente coincide con la definición sintáctica de las leyes científicas (Carnap 1966); la diferencia, no obstante, está presente en su componente *procedural*: las reglas de producción son componentes programables de la teoría, de modo que las asociaciones entre conceptos, aunque expresadas verbalmente, no están sometidas a un único mecanismo de producción (Holland et al. 1986).

En el segundo caso (representación topológica de cadenas de inferencias) utilizaremos unos sistemas de programación denominados *Redes Neuronales* (cf. Aleksander y Morton 1990, Caudill y Butler 1989). Estos sistemas están constituidos por un gran número de elementos simples de procesamiento conectados entre sí. Los elementos (“neuronas”) no contienen conocimiento declarativo, sin embargo las conexiones que los unen tienen asociadas un valor numérico: el peso o la intensidad que determinan su importancia e influencia en la red neuronal. Ese valor numérico se define como medida de la frecuencia con que la neurona de destino y la neurona de origen han estado excitadas simultáneamente. Ese esquema de conexiones o *vector de excitaciones* permite almacenar la información utilizando como símbolos los pesos o estimaciones numéricas de las conexiones.

En ambos casos, las inferencias se representan por medio de las asociaciones entre dos o más unidades de conocimiento, lo que permite que sus contenidos respectivos (ya sea un valor proposicional o la estimación del peso de una conexión) pase de la una a la otra. Este tipo de inferencia suele denominarse *abducción* (Holland et al. 1986, Thagard 1988, O'Rourke et al. 1990). El más sencillo de los operadores abductivos nos servirá para caracterizar mejor esta modalidad de razonamiento. Supongamos que deseamos saber por qué un artefacto *a* tiene la propiedad *P*; disponemos de cierta información previa (todos los *F* son *P*), por lo tanto estableceremos una asociación abductiva entre “*a* tiene *P*” y “*a* es *F*”, y utilizaremos la segunda unidad de conocimiento (*a* es *F*) para *explicar* por qué *a* tiene *P*. Observe que la conexión entre ambas unidades no es segura; la hemos *conjeturado* sobre la base de un conocimiento general previo (“todos los *F* son *P*”); es decir, el razonamiento abductivo propone ciertas hipótesis plausibles acerca de una situación inicial, con el fin de explicar unas observaciones. Josephson et al. (1987) sugieren adoptar el siguiente pseudo-silogismo para representar las inferencias abductivas:

D es una colección de datos (observaciones, descripciones, imágenes digitalizadas, etc.)

H explica D (Si H fuese verdadero, entonces *implicaría* a D)

Ninguna de las Hipótesis conocidas explica D mejor que H.

Entonces, H es correcto.

El operador abductivo por excelencia es la *analogía* (cf. Thagard 1988, Falkenhainer 1990), aunque los lógicos señalan también el interés de los siguientes: igualdad, simetría, homología, desigualdad y referencia. Así, dos estados de un problema estarán asociados si dos estados “semejantes” lo están (definición de Analogía, cf. Owen 1991), o bien, si un estado intermedio hace referencia a otro en su representación, estableceremos una asociación entre ambos, etc.

Por consiguiente, dos unidades abductivamente asociadas son dos unidades entre las cuales se ha establecido una relación heurística; es decir, una relación que no está basada en la naturaleza profunda de las unidades, sino en un criterio externo establecido por el investigador (Clancey 1984). Más de un lector preguntará por qué son tan débiles, formalmente hablando, las asociaciones que configuran una cadena de inferencias. La razón estriba en las características del mecanismo *heurístico* de producción de interpretaciones desarrollado en la sección anterior. La operación de búsqueda de la “mejor solución posible” no puede implementarse mediante un algoritmo de tipo polinomial, es decir, *no existe ningún procedimiento que nos permita averiguar, en un número finito de pasos, la idoneidad de una interpretación dada cierta información empírica inicial* .

Tanto los especialistas en lógica como en informática le han dado muchas vueltas a este corolario de los Teoremas de Gödel y de Church, y han llegado a la conclusión que no hay una respuesta *formal* al mismo, aunque sí la hay *heurística*: el mecanismo de búsqueda heurística debe implementarse en términos de un conjunto finito de operaciones independientes, cada una de ellas con sus propios criterios heurísticos de validez. Ese mecanismo será muy diferente según utilicemos un modelo proposicional de representación o un modelo topológico.

En el primer caso, el modelo proposicional de representación, el mecanismo de razonamiento que utilizaremos para generar una interpretación arqueológica adopta el aspecto de una secuencia de operaciones lógicas muy sencillas, ninguna de las cuales por sí misma, es lo suficientemente sofisticada para generar por sí sola la interpretación, si bien, el esfuerzo combinado de todas esas operaciones permite construir un estado final aceptable. Es decir, un gran número de operadores abductivos muy simples “activa” las distintas unidades de conocimiento que configuran la cadena de inferencias. Dado que en esa cadena las unidades son muy distintas entre sí, las “condiciones de activación” diferirán entre unas y otras, teniendo en cuenta, además, el orden o situación de cada unidad en la cadena de inferencias. Se necesita, por tanto, una “reacción en cadena” para poder “activar” la interpretación, ya que estamos obligados a usar como conocimiento previo las unidades de conocimiento que han sido activadas previamente. En términos computacionales, esa reacción en cadena es el resultado de la ejecución de un mecanismo de *propagación de la función de activación*. Ese mecanismo puede llegar a ser bastante complejo, computacionalmente hablando, ya que depende del *orden* de las unidades en la cadena de inferencia –el cual depende, a su vez, de la naturaleza de las diferencias entre las distintas unidades–, así como de la naturaleza de la función de activación de cada una de ellas; por lo general se trata de una

función cualitativa y, por lo tanto, no continua, que enumera los estados intermedios del problema “activados” entre el estado inicial y el estado final

En el caso de la *representación topológica de las inferencias*, las asociaciones entre unidades adoptan otro aspecto, lo que condiciona el tipo de cadena de inferencias obtenido. La excitación (equivalente a la “activación” proposicional) de uno de los elementos de procesamiento (“neurona”) que configuran la red neuronal, es el resultado de la excitación de un número x de neuronas anteriores, así como de la existencia de conexiones excitatorias (conexiones con un peso elevado) entre unas y otras. Ahora bien, esa función –continua y no lineal, a diferencia de la “activación” proposicional– no transfiere información declarativa: el valor excitatorio de las neuronas y de sus conexiones no tiene sentido fuera del entorno de programación. Por consiguiente, las asociaciones abductivas no tendrán que buscarse entre las neuronas, sino entre los *vectores de excitación*. Por ejemplo, para interpretar neuronalmente la función de un útil lítico, por ejemplo, tenemos que traducir previamente la asociación proposicional –más intuitiva–

(Rasgos descriptivos (x), (y), (z)) → Raspador Carenado

en un vector de elementos excitados y pesos de sus conexiones respectivas. Ello se realiza mediante ciertas ecuaciones que tienen en cuenta el número de ejemplos y de rasgos que describen los raspadores carenados conocidos. Siempre que introduzcamos la misma lista de rasgos descriptivos, “excitaremos” la misma interpretación, que ha sido “aprendida” por el programa a partir del análisis de un determinado número de ejemplos de raspadores carenados. Observese que se ha establecido una relación de analogía entre el conjunto de rasgos descriptivos (estímulo) y la respuesta generada por el sistema. Que esa “respuesta” esté representada *topológicamente* (en términos de un vector de excitaciones) no afecta al contenido de la analogía. La cadena de inferencias, en este caso, tiene tan sólo un estado inicial y un estado final, sin estados intermedios. El inconveniente es que la interpretación ha tenido lugar en una *caja negra*: A no ser que el investigador sea un matemático avezado, no podrá *visualizar* el mecanismo que ha establecido las asociaciones.

Los enlaces o asociaciones que definen una cadena de inferencia son, en realidad, unidades de conocimiento en sí mismas. Su significado suele definirse en términos “explicativos” (Holland et al. 1986, Thagard 1988). Es decir, dos unidades de conocimiento están asociadas cuando una de ellas *explica* a la otra. Tradicionalmente la definición rigurosa del término *explicación* se ha resistido a los intentos de la mayoría de filósofos de la ciencia. Si seguimos a un autor como Paul Thagard (1989) tendremos que admitir que lo que nos ha conducido a establecer una “conexión” explicativa son motivos meramente prácticos: A explica B porque A cumple ciertos requisitos. Esos requisitos pueden ser muy diversos, formales o quasi formales o bien derivados del objetivo formulado al enunciar el problema interpretativo (en ese caso lo denominaremos requisito heurístico). Lo cierto es que no existe una definición formal (y, por tanto, computable) del término *explicación*. En ausencia de una definición operativa debemos trabajar bajo el supuesto –probablemente erróneo– que una explicación equivale a una asociación objetiva, en donde el adjetivo “objetivo” no hace referencia a ninguna verdad universal, sino tan sólo al ajuste entre las condiciones de activación de una unidad de conocimiento y el objetivo explícito con el que se formula el problema.

DEDUCCIÓN E INDUCCIÓN EN EL RAZONAMIENTO ARQUEOLÓGICO

Hasta no hace mucho se creía que una cadena de inferencias estaba constituida, necesariamente, por unidades de conocimiento asociadas *deductivamente* (Watson et al., 1971). Uno de los resultados de la arqueología automática radica, precisamente, en señalar la naturaleza *heurística*, y por tanto, no-deductiva, de las asociaciones que configuran la cadena de inferencias. Ello no significa, sin embargo, que las deducciones no tengan utilidad en el razonamiento arqueológico.

Los investigadores en Inteligencia Artificial y Psicología Cognitiva han contribuido a redefinir lo que la Lógica Clásica entendía por deducción (cf. Blasius y Burckert 1989, Johnson-Laird y Byrne 1991). Un razonamiento deductivo correcto es aquel cuya conclusión es verdadera en cualquiera de las situaciones en las que la premisa es verdadera; por tanto, denominaremos *deducción* a una asociación entre dos unidades de conocimiento tal que estemos seguros que la segunda de ellas es una consecuencia de la primera. En otras palabras, entre ambas unidades existe una relación *necesaria* que nace en la propia definición de las unidades y no de una condición heurística externa impuesta por el investigador.

Para poder establecer esa relación de *consecuencia* entre ambas unidades, es preciso que la segunda de ellas (la unidad *deducida*) contenga explícitamente cierta información que ya existía, aunque de forma implícita, en la primera de ellas (la premisa). A diferencia del razonamiento abductivo, el razonamiento deductivo es incapaz de *crear* nueva información, tan sólo vuelve explícita lo que antes se conocía de forma implícita.

Es fácil reproducir este tipo de razonamiento en un ordenador; basta con que implementemos el mecanismo lógico denominado *modus ponens* :

Si

El Estado inicial del problema que se caracteriza por la presencia de los rasgos descriptivos (x) (y) (z) está *siempre* asociado a un Estado Final caracterizado por los rasgos (a) (b) (c)

Entonces,

Dado cualquier estado inicial que contenga los rasgos (x)(y)(z)

Deduciremos,

La interpretación de (x)(y)(z) es (a)(b)(c).

Observe que para establecer una asociación deductiva es preciso que esa asociación haya sido validada previamente. No basta con que conjeturemos heurísticamente la asociación, ni que creamos que es la “mejor” de todas las posibles; hemos de estar seguros que en ningún caso encontraremos una interpretación mejor. Para estar seguros de una afirmación es necesario que demos la verdad de cada una de las instancias de esa información. ¡Pero puede haber infinitas instancias! No hay ordenador ni cerebro humano capaz de analizar uno por uno todos los contextos posibles en los que determinada información deba ser, necesariamente, verdadera. Por tal motivo, tanto lógicos como informáticos recurren a ciertos principios que reduzcan de algún modo la evaluación de la “verdad” de una asociación deductiva:

- recurriendo a reglas formales que validen sintácticamente la validez de la asociación. Siempre que el razonamiento reproduzca esa sintaxis lo consideraremos “verdadero”.

- recurriendo a reglas semánticas fijas, esto es, reglas cuyo contenido sabemos válido. Se trata de una variante del *modus ponens*.

- recurriendo a una simulación del contenido de la asociación (Johnson-Laird 1983):

- se construye un modelo explícito del conocimiento contenido en las premisas

- se construye un modelo explícito del conocimiento en el consecuente.

- se intenta unir ambos modelos, analizando su compatibilidad (ausencia de contradicciones).

La aplicación de alguno de estos mecanismos de asociación nos permite establecer, automáticamente y sin necesidad de pruebas ulteriores, la validez e idoneidad de la interpretación generada. Ahora bien, si este mecanismo de asociación fuese el único modo de combinar información, la mayor parte de problemas arqueológicos quedaría sin resolver. Si los conceptos asociados en una cadena de inferencias no cumplen las condiciones siguientes:

- la aplicabilidad de la conclusión ha de ser mayor o igual que la aplicabilidad de las premisas de las que ha sido deducida, nunca menor.

- la conclusión no puede afirmar algo que se haya afirmado con anterioridad.

- la conclusión ha de afirmar algo que no esté explícito (aunque necesariamente, debe estar implícito) en el enunciado de las premisas.

diremos que de la información empírica inicial no puede *deducirse* ninguna interpretación por falta de conocimiento fiable (experimentalmente válido). Ello no quiere decir que ese estado inicial no sea interpretable, sino que la validez de la interpretación (obtenida gracias a los operadores abductivos) no se deriva *exclusivamente* de la descripción inicial. Tal y como hemos visto en la sección anterior, las inferencias no-deductivas (abductivas) son mucho más frecuentes.

En ausencia de enlaces deductivos que garanticen formalmente que cada una de las unidades de conocimiento que aparecen en la cadena de inferencias ocupa el lugar que le corresponde, tendremos que utilizar un criterio heurístico de validación: la interpretación generada tiene que ajustarse a las especificaciones del objetivo del problema. Los *objetivos* de un problema suelen representarse por medio de proposiciones que hay que validar, es decir, por medio de preguntas:

¿(x) se fecha en el siglo IX a.C.?

Ahora bien, tan sólo en el caso en que dispongamos de una gran cantidad de conocimiento bien estructurado podremos resolver el problema respondiendo a una sola pregunta. En la mayoría de las

ocasiones, es preciso que hagamos una gran cantidad de preguntas para poder encontrar la solución. Cuando se le plantea una pregunta al sistema, éste activa una de las unidades de conocimiento utilizando alguno de los operadores asociados específicamente a esa unidad. Si el sistema no puede decidir qué operador es el que hay que utilizar, deberemos plantear otra pregunta de tal modo que la respuesta a la misma proporcione la información necesaria para activar la solución a la primera que hemos formulado.

En otras palabras, los “fallos” en la resolución dan lugar a que nos planteemos *subproblemas* cuyo objetivo es, precisamente, resolver la nueva dificultad aparecida reduciendo la distancia entre el estado inicial y alguna de las soluciones. Por consiguiente, para poder generar una interpretación precisamos de un *plan* o secuencia de subproblemas. La apariencia externa de ese plan es la de una compleja red en la que están integrados todos los subobjetivos, unidos entre sí por medio de relaciones de “cooperación”, pues cada subproblema, con ayuda de los demás contribuye a encontrar la, solución idónea al estado inicial.

Los subobjetivos son unidades de conocimiento preespecificadas cuyo objetivo es, precisamente, imponer un orden determinado en el flujo de asociaciones, esto es, en la propagación de la función de activación. Se trata de un conocimiento acerca del procedimiento para resolver el problema, antes que acerca de la solución correcta del problema; nos referiremos a él como *meta-conocimiento*.

La secuencia de subobjetivos ha de estar deductivamente ordenada, de tal modo que un subobjetivo debe deducirse del anterior. Es precisamente esa secuencia *deductiva* la que nos permitirá validar la cadena de inferencias que hemos obtenido abductivamente: cuando la cadena de inferencias reproduzca la secuencia de subobjetivos, el estado final de la misma se ajustará con lo especificado en el objetivo del problema. Aquí radica la noción de *objetividad* (ajuste con los objetivos) y no en ninguna esotérica verdad universal inmanente.

Todas estas dificultades son exclusivas del modelo proposicional de representación de las cadenas de inferencia. Muchos autores piensan que este modelo está limitado por culpa de la necesaria unidimensionalidad del esquema de asociaciones. Sin embargo, hemos visto en la sección anterior que para interpretar los datos arqueológicos podemos (y en muchos casos debemos) utilizar cualquier combinación de enlaces asociativos. Es precisamente la diversidad de los operadores asociativos necesarios para configurar una interpretación la que debiera prevenirnos antes de encerrarnos en cadenas de inferencia con una estructura predefinida.

¿Por qué limitarnos entonces al uso de estructuras de control de naturaleza deductiva, mucho más exigentes, formalmente, y que exigen grandes cantidades de conocimiento válido, del cual carecemos en la mayoría de las ocasiones? Porque deseamos estar razonablemente seguros que la interpretación es adecuada, y nadie ha dicho que el razonamiento científico tenga que ser sencillo. Ahora bien, ¿no existirá algún método de validación que no exija tanto *meta-conocimiento*? La respuesta es negativa, dentro del modelo proposicional de representación, pero afirmativa dentro del modelo topológico.

La combinación de enlaces asociativos diversos debiera dar lugar a una estructura en paralelo o reticular y no a una cadena o secuencia de unidades. En esa estructura la función de activación no se propagaría unidimensionalmente, sino que se distribuiría en infinitas direcciones. Obviamente, esas estructuras multidimensionales han de ser representadas mediante sistemas topológicos, esto es, *vectores de excitación*. Quizás el formato de las redes neuronales, tal y como ha sido expuesto en la sección anterior no sea el más conveniente, dada su poca transparencia (Partridge 1990), de ahí el modelo mixto diseñado por algunos autores, en el que las “neuronas” tienen contenido declarativo y las conexiones excitatorias representan la influencia causal de las distintas unidades de conocimiento (Feldman y Ballard 1982, Shastri

1988). Se trata de un sistema quasi-proposicional, en el que el orden de las inferencias puede establecerse matemáticamente a partir del peso de los enlaces entre las neuronas (cf. Churchland 1989).

En tales circunstancias, el mejor modo de calcular el ajuste entre la interpretación y el objetivo es por medio de operaciones de *optimización*, esto es, el cálculo de la mejor de todas las funciones continuas que relacionan el estado inicial con el final. No detallaremos aquí las ecuaciones necesarias para *optimizar* una interpretación (cf. Pao 1989, Zeidenberg 1990), si bien son necesarias ciertas apreciaciones: algunos algoritmos de optimización (por ejemplo el algoritmo de Kohonen) utiliza criterios formales muy estrictos, según los cuales la mejor solución es aquella que adopta una estructura matemática determinada, usualmente basada en Teoría de las Probabilidades. Obviamente esa solución nos devuelve a la polémica positivista acerca del uso de reglas sintácticas para demostrar la verdad de las afirmaciones científicas. Por otro lado, los algoritmos más comunmente usados en la programación de redes neuronales (el algoritmo de retro-propagación) exige también meta-conocimiento. No obstante, ese “conocimiento acerca del problema” es muy distinto del que necesitábamos en el caso proposicional: en lugar de desarrollar el plan del problema (secuencia de sub-objetivos), se precisa de un conjunto adecuadamente grande de ejemplos a partir de los cuales puedan calcularse las propiedades topológicas de la solución ideal.

Hasta ahora no hemos mencionado para nada el razonamiento *inductivo*, fundamental para muchos arqueólogos (Clarke 1968): los conceptos científicos, las interpretaciones, se generaban por medio de ciertos operadores estadísticos (correlación de atributos), utilizando como información inicial una descripción de los datos. Según este enfoque, las interpretaciones se expresarían mediante series de atributos estrechamente correlacionados y el mecanismo de producción de *interpretaciones* se reduciría a la aplicación de operadores universales sobre la información inicial, sin que se precise ningún tipo de hipótesis previa cuya relación con la evidencia fuese preciso establecer.

Las investigaciones recientes en Inteligencia Artificial y en Psicología Cognitiva (Holland et al., 1986, Langley et al., 1987) han redefinido esa modalidad de inferencia. En esos trabajos puede leerse una concepción muy distinta de la filosofía inductiva y de su relevancia en la investigación científica, muy en línea con la revisión de la filosofía de Francis Bacon (1561-1626) emprendida por Peter Urbach (Urbach 1982, 1987, Gibbins 1990). Según este autor, las críticas vertidas a Bacon por filósofos posteriores son infundadas. El Francis Bacon descrito por Urbach se parece curiosamente a Popper: aunque Bacon oponía la *interpretatio naturae* a la *anticipatio mentis*, nunca se opuso al razonamiento por medio de hipótesis, sino que se oponía al predominio de las hipótesis previas sobre la investigación empírica. En otras palabras: el razonamiento por hipótesis no debía sustituir al análisis empírico (búsqueda de regularidades).

Los arqueólogos no analizan los fenómenos sociales aplicando fórmulas mágicas a sus datos y usando el resultado de esas fórmulas como conceptos o leyes absolutas. Debieran ser capaces de *aprender conceptos* y no tan sólo señalar la existencia de co-ocurrencias. ¿Es posible “aprender” un concepto nuevo describiendo exclusivamente la regularidad implícita en la evidencia? No, porque tanto los informáticos como los lógicos nos han enseñado que los conceptos científicos no pueden ser generados algorítmicamente. Hemos de tener en cuenta, además, que la noción de “aprendizaje” suele definirse como “aquel fenómeno exhibido cuando un sistema (agente humano o máquina) mejora en la ejecución de una tarea concreta, sin necesidad de haber sido reprogramado” (Anderson 1983, Newell 1990). Por consiguiente, la tarea del arqueólogo es *modificar* progresivamente la formulación de un concepto inicial (hipótesis), acercándose paulatinamente a un concepto “mejor” que el de partida. “Mejor” puede

significar cosas distintas: un mayor porcentaje de soluciones correctas, una respuesta más rápida, soluciones más prácticas o eficaces, o con un mayor abanico de posibilidades de aplicación, un ajuste más apropiado con el objetivo, etc.

El aprendizaje de una categoría suele definirse en Psicología Cognitiva en tanto que *transferencia* de conocimiento de un conjunto de ejemplos a una expresión lógica. Sin embargo, no toda transferencia de conocimiento permite aprender; es preciso que el conocimiento de la segunda unidad *augmente* como consecuencia de esa transferencia. Pero no hay aprendizaje, no hay “incremento” del contenido, si no se definen previamente los parámetros que permitan establecer la “mejora” del sistema. Aquí radica el *problema de la inducción*, que durante siglos se han planteado los filósofos: a diferencia de lo que sucede con la deducción, no puede probarse si las leyes o entidades generales resultantes de una inferencia inductiva son correctas. Mientras que la deducción es un tipo de razonamiento finito, es decir, concluye tras un número determinado de pasos, la inducción no concluye nunca, porque no puede establecerse con seguridad cual de sus resultados es el definitivo. De este modo, la creación inductiva de conceptos científicos es en realidad un proceso continuo, y no se reduce a la aplicación de un único operador. Por consiguiente, lo realmente importante en una inferencia inductiva no es la naturaleza de los operadores (estadísticos o lógicos), sino las heurísticas de procesamiento que asegurarán que las conclusiones inductivas son plausibles y relevantes para el objetivo que se está procesando.

En definitiva, el control de la inducción tiene más importancia que la inducción en sí misma. Hay que limitar la búsqueda de generalizaciones y/o asociaciones, aun con el riesgo de omitir informaciones importantes. No puede haber aprendizaje si no se controla de un modo u otro el mecanismo de inducción, ya se trate de la generalización o de la correlación de propiedades. Cuanto más estrictas sean las condiciones que imponamos al mecanismo, mayor riesgo de eliminar generalizaciones interesantes; si esas condiciones fuesen demasiado débiles, las dimensiones del espacio de búsqueda serán excesivas para encontrar en él hipótesis satisfactorias.

De este modo, en el dominio de la Inteligencia Artificial, las inferencias inductivas se caracterizan de un modo bastante distinto al de la inferencia estadística clásica: el objetivo no es distinguir qué atributos o variables están correlacionados con qué objetos en una base de datos, sino en modificar progresivamente la definición de un objeto paradigmático de modo que la definición final sea utilizable para todas las instancias posibles del mismo objeto. Se trata de un tipo de inferencia que ejecutaremos en respuesta a ciertos objetivos específicos, tales como la búsqueda de una explicación plausible para un resultado inesperado. *El estudio de la inducción es, pues, el estudio de cómo se modifica el conocimiento a través de su uso.*

A diferencia de lo que creían los primeros partidarios de la nueva Arqueología no necesitamos mecanismos formales estrictos (deducción lógica) para interpretar los datos arqueológicos, aunque nada nos impida utilizarlos en la resolución de determinados problemas. Ahora bien, que la deducción no sea el mecanismo fundamental no significa que no existan otros mecanismos, ni que la arqueología no pueda ser “racional”. Lo que denominamos “razonamiento arqueológico” no es más que un conjunto de mecanismos de *combinación conceptual* (Thagard 1988, 1990). El investigador debe combinar la información que dispone para poder producir interpretaciones (conceptos). Esa combinación, sin embargo, no es una operación formal; existen numerosísimos operadores combinativos, cada uno de los cuales será útil en determinadas circunstancias. El único requisito “científico” para que una combinación conceptual produzca una inferencia válida es que la secuencia de combinaciones se haya producido siguiendo un plan específico y no al azar. En otras palabras, ha de resultar posible, en todo momento,

saber *cómo* (secuencia de combinaciones) y *por qué* (plan que hay que seguir para resolver el problema) se ha generado esa interpretación y no otra.

LA NATURALEZA DE LAS TEORÍAS ARQUEOLÓGICAS

En las secciones anteriores hemos podido apreciar la necesidad imperativa de conocimientos *estructurados* para poder resolver problemas arqueológicos. Los arqueólogos no debieran limitarse a la mera acumulación de datos empíricos simples, sino que deben estructurar y relacionar toda la información que consiguen para poder llegar a resolver nuevos problemas interpretativos. La tarea del arqueólogo es *asociar* unidades de conocimiento; partiendo de asociaciones inductivas producirá, paulatinamente, una jerarquía de entidades teóricas cada vez más alejadas del nivel observacional y cuya generación requerirá cadenas de inferencia extraordinariamente largas y una enorme variedad de operadores heurísticos. Una vez validadas esas asociaciones, se convertirán en unidades de conocimiento de pleno derecho y pasarán a formar parte del conocimiento preexistente, a la espera de ser asociadas a otras unidades. En definitiva, una Teoría contiene interpretaciones que han sido generadas, en algún momento anterior como solución a cierto problema.

Esta importancia de lo *asociativo* afecta también al significado o identidad semántica de las unidades de conocimiento: el significado de las entidades teóricas no deriva de su aplicación observacional, sino que depende de las inferencias (cadenas de asociaciones) en las que toma parte. En otras palabras, la identidad semántica de un concepto está determinada por su situación peculiar en la red de asociaciones (cf. Churchland 1989).

Si queremos “programar” en un ordenador una Teoría Arqueológica, habremos de tener en cuenta todas esas consideraciones. Una teoría no es una base de datos, sino un programa aparentemente “inteligente” capaz de operar con nuestros datos, permitiendo inferencias automáticas, esto es, la combinación de datos e hipótesis para generar nuevos datos que resolvieran los problemas que tenemos planteados. En cualquier caso, una *Teoría Computacional* puede describirse en tanto que base de datos con la propiedad de “modificarse” a sí misma para realizar un fin concreto, pues, sin finalidad, no habría “inteligencia”. En cierto sentido podríamos decir que una teoría computacional es un programa de ordenador *capaz de utilizarse a sí mismo con un objetivo concreto*; en nuestro caso, ese objetivo consistirá en resolver un problema arqueológico.

La utilización de una teoría computacional se diferencia, pues, del empleo de bases de datos clásicas (relacionales). En estas últimas de lo que se trata es de “recuperar” ciertas unidades de conocimiento (los datos) que se han introducido previamente: el usuario *consulta* la información. Para llevar a cabo esta tarea, el programa gestor de la base de datos dispone de un lenguaje especializado en el que se programarán las consultas, las cuales estarán representadas por medio de una enumeración simple de las características que ha de cumplir el dato buscado. El usuario de la teoría computacional, por el contrario, no *consulta* información, sino que “activa” aquellas unidades de conocimiento que el programa considera son las soluciones adecuadas al problema planteado. Puede pensarse que esa “activación” no deja de ser una consulta maquillada; sin embargo hay una diferencia fundamental entre la interrogación de una base de datos clásica y la interrogación de una teoría computacional: se necesita *información contextual* para activar las posibles soluciones al problema. Es decir, los contenidos de la teoría computacional serán activados sí y sólo si son relevantes en el contexto marcado por el problema que hay que resolver.

Aún no se ha podido programar una teoría computacional en su integridad, aunque sí se conocen los componentes informáticos que deberá tener: un mecanismo de solución de problemas capaz de tratar inferencias para construir la estrategia de satisfacción de ciertos objetivos, una memoria accesible (asociativa), y un mecanismo de inducción que genera nuevo conocimiento y lo conserva en la memoria. Es importante señalar que, en esta estructura, los objetivos, y por consiguiente la estructural de control deductivo del flujo de inferencias, no pertenecen al sistema cognitivo, sino que son introducidos por el usuario siempre que éste utilice los contenidos de la teoría. Es obvio que el investigador propone esos objetivos, en tanto que miembro de un determinado grupo social; el razonamiento arqueológico está afectado por los valores sociales. Sin embargo, en tanto que conjunto *artificial* de mecanismos que depende de la voluntad del investigador, nada impide trascender esos valores. Todo depende de los riesgos que el investigador desee asumir.

En definitiva, una teoría computacional debe contener:

- un conjunto de conceptos que describan el conocimiento relevante al problema que se quiere resolver.

- una representación adecuada de esos conceptos, esto es, una representación *activa* que permita que los conceptos reaccionen a los mensajes enviados por el usuario o por los otros conceptos del Sistema.

- un conjunto de reglas que gobiernen la descripción de los conceptos en términos de su representación.

- un conjunto de operadores que actúen sobre las representaciones.

Otra de las características básicas de cualquier teoría computacional es que su arquitectura interna sea análoga al mecanismo de razonamiento que hemos descrito en páginas anteriores, puesto que, al margen de interpretar las observaciones empíricas, la teoría debe proporcionar una explicación de las unidades de conocimiento usadas para explicar las observaciones; es decir, debe especificar el *cómo* y el *por qué* de la interpretación generada. Según este principio, la arquitectura idónea de una Teoría Computacional sería modular (cf. Pitrat 1990):

- 1) para responder al objetivo F, se necesita el conocimiento C
- 2) el objetivo G hace referencia cómo se ha resuelto E con ayuda de C
- 3) para responder al objetivo G se necesita el conocimiento D
- 4) el objetivo H hace referencia a la relación entre C y D
- 5) para responder al objetivo H se necesita el conocimiento E

y así *ad infinitum*.

Una consecuencia importante de este enfoque estriba en la necesidad de diferenciar claramente la tarea de adquirir el conocimiento necesario para resolver un problema situado más arriba en la escala de abstracción que acabamos de presentar. Un programa que carezca del conocimiento D, por ejemplo, será incapaz de resolver el problema G, pero no por ello le faltará “inteligencia”: es inteligente con respecto al problema F, aunque no lo sea con respecto al problema G. Ningún filósofo de la ciencia debiera de extrañarse; la respuesta a la paradoja de Russell y al Teorema de Gödel pasa por construir sistemas formales que incluyan a un sistema formal incompleto. Como *todos* los sistemas formales son incompletos por definición, la tarea de construir sistemas que los incluyan no acabará nunca; tampoco el razonamiento científico tiene un final, aunque el conocimiento sea acumulable.

Los datos y conceptos científicos son, esencialmente, dinámicos, de ahí que la Teoría Computacional deba representar de un modo u otro, el cambio entre dos momentos o estados. Muchos avances científicos dependen de la introducción de nuevos conceptos que sólo al final demuestran no ser válidos. Su indeterminación suele descubrirse a la larga, porque ha aparecido una situación inesperada en la que un conocimiento, hasta entonces válido, genera conflictos. En otros casos, el investigador es consciente que está trabajando con una información disyuntiva, incompleta, negativa o meramente implícita. De ahí que el razonamiento científico precise de la facultad de elaborar inferencias plausibles en presencia de una información que a menudo es incompleta y evolutiva, esto es, cambiante con el tiempo. Este peculiar tipo de razonamiento no se deja formalizar por la lógica clásica, que trabaja con premisas verdaderas universalmente, completas y no cambiantes. En el razonamiento científico, las conclusiones obtenidas en un momento dado, ya no serán válidas cuando se haya revisado la teoría que la produjo. En otras palabras, la principal característica del razonamiento científico es la de basarse en inferencias “revisables”, o no verdaderas tautológicamente.

Las teorías computacionales han de permitir inferencias revisables; para ello tendremos que implementar las dos propiedades fundamentales de ese tipo de razonamiento, en primer lugar la propiedad de *no-monotonía*, que se define del siguiente modo:

lo que es deducible en el instante t , no lo es necesariamente en el instante $t + 1$.

En otras palabras, el número de conclusiones válidas (extensiones) en una inferencia científica, puede decrecer al aumentar las premisas de la misma.

La otra propiedad del razonamiento “revisable” es la *pluri-extensionalidad*: a partir de un mismo conjunto de informaciones iniciales, pueden obtenerse distintos conjuntos incompatibles de conclusiones. Es decir, dado un mismo problema y una única Teoría, pueden inferirse distintas soluciones. Ya hemos visto que se trata de una consecuencia del uso de operadores heurísticos “débiles” y de condiciones heurísticas de validación, así como de la no-monotonía del razonamiento. Un cierto grado de pluri-extensionalidad es aconsejable para eliminar toda tentación *formalista* en el diseño de una teoría computacional, sin embargo, el riesgo de incurrir en múltiples interpretaciones incoherentes entre sí es demasiado alto. La programación de sistemas no-monótonos en los que la pluri-extensionalidad no degenera en contradicciones constituye uno de los dominios más dinámicos de la Inteligencia Artificial. (cf Besnard 1989, Grégoire 1990, Lukaszewicz 1990).

La programación de Teorías Computacionales nos permite apreciar que es la teoría la que proporciona las respuestas a un problema interpretativo, y no el teórico. Tanto los seres humanos como las teorías son agentes “racionales” capaces de responder preguntas. Los seres humanos, sin embargo, no

suelen proporcionar información adicional acerca de la corrección de su respuesta, y si lo hacen están obligados a recurrir a una teoría. En ausencia de criterios de validación externos, los seres humanos emplean la argumentación como mecanismo de crítica y reformulación, evaluando la idoneidad de las respuestas con arreglo al cumplimiento o no de sus propios objetivos individuales (subjetivos), prescindiendo de los objetivos sobre los que estaba basada la pregunta. Por el contrario, una teoría científica – computacional o no– es un cuerpo *explícito* de conocimiento, a partir del cual cualquier persona lo suficientemente entrenada podrá extraer respuestas a sus preguntas.

Obviamente, qué preguntas encontrarán respuesta y qué preguntas permanecerán sin solución dependerá de la preparación del usuario de la teoría; pero no así la respuesta en sí, la cual, una vez obtenida, depende exclusivamente de la teoría; en nuestro caso, del conocimiento implementado en el ordenador y de la manera en que ha sido implementado.

ARQUEOLOGÍA AUTOMÁTICA

Si queremos saber cómo funciona la mente, debemos preguntarnos qué es lo que hace y por qué hace lo que hace.

Esta cita es fácilmente aplicable al ámbito de la investigación arqueológica: si deseamos saber qué es una interpretación o una teoría debemos averiguar qué es lo que realmente hacen los arqueólogos cuando “razonan”, esto es, cuando interpretan las evidencias arqueológicas. En Ciencias Cognitivas ha adquirido una relativa fortuna la distinción entre una Teoría en el nivel computacional y una teoría en el nivel algorítmico (Marr 1982). En su nivel computacional, una teoría está caracterizada por aquello que se computa, por qué está siendo computado y a qué condiciones debe ajustarse el procedimiento. Una teoría en el nivel algorítmico especifica cómo se realiza la computación, y debiera ser lo suficientemente precisa como para poder ser simulada mediante un ordenador. Según Chomsky (1965) una teoría en el nivel computacional es una explicación de la competencia humana, la capacidad del investigador para resolver sus problemas, en definitiva, sus *objetivos*. En su realización algorítmica, la teoría debe explicar las características de ejecución humana, dónde falla y por qué nos conduce, en ocasiones, al error; debiera facilitar, igualmente, su integración con las restantes acciones cognitivas.

La Arqueología Automática es, obviamente, una teoría del razonamiento arqueológico en su nivel algorítmico. Su principal objetivo es el análisis de la racionalidad de una teoría por medios computacionales. En otras palabras, averiguar el significado de las interpretaciones arqueológicas en términos de los mecanismos que permitieron construir esa interpretación y que determinan su uso durante la resolución de un problema. (Thagard 1988).

El estudio de esa racionalidad suele emprenderse distinguiendo en la Teoría un *componente lógico* (sintáctico) de un *componente subjetivo* (semántico). El primero está incorporado a cualquier programa de ordenador (depurador), y su función es determinar la gama de inferencias incompatibles, lo cual debiera preservar la consistencia del conocimiento arqueológico. El segundo determinará la efectividad de las inferencias. Ahora bien, hemos visto que la efectividad de las soluciones propuestas por la Teoría computacional sólo puede realizarse *empíricamente*. La propiedad de la pluri-extensionalidad del razonamiento científico impone una serie de condiciones y límites difícilmente formalizables a la selección del “mejor” conjunto de soluciones posibles. En general, el criterio de selección parece estar

definido en los objetivos iniciales del problema a resolver, por lo que el objetivo final de cualquier análisis de la racionalidad de una Teoría parte del principio según el cual la calidad de una Teoría Computacional no sólo depende de la calidad del conocimiento que contiene, sino del uso al que se destina esa Teoría.

El *componente subjetivo* (semántico) de las Teorías Computacionales representa al usuario o al acto de *utilizar* el conocimiento. Algunos autores prefieren dejar sin programar ese *componente subjetivo*. La Teoría Computacional asegura tan sólo el componente lógico (sintáctico) de la *racionalidad* de una investigación; de lo que se servirá el investigador como instrumento auxiliar en la resolución de un problema. En estas circunstancias, la evaluación final de la Teoría y la validación de la solución obtenida corresponderá al usuario. Este hecho ha llevado a muchos programadores a enfatizar la relevancia del diálogo o comunicación interactiva Máquina/Usuario, antes que los mecanismos automáticos de revisión, limitados por definición, ya que dependen del metaconocimiento que se haya introducido en el sistema.

Sin embargo, nada nos impide llevar más lejos el análisis computacional de la racionalidad de una Teoría y programar una Teoría Computacional cuyo dominio de aplicación sea, precisamente, el *uso* de una Teoría anterior, para lo cual simulará la actuación de un usuario hipotético. De este modo implementaremos el componente semántico de la racionalidad de la Teoría. Conviene tener presente que el *uso* de la Teoría, esto es, la resolución de un determinado problema científico debe ser consistente con los objetivos marcados, objetivos que constituyen las unidades de conocimiento características de la *meta-teoría*.

En definitiva, el propósito de estos programas de ordenador no es sustituir a los arqueólogos por máquinas inteligentes, sino estudiar los mecanismos de resolución de problemas utilizados por los científicos sociales, o los mecanismos de resolución que *debieran* utilizar. Automatizando el procedimiento de interpretación del registro arqueológico descubriremos los mecanismos cognitivos subyacentes en todo razonamiento científico. La ventaja de la automatización y del empleo de las técnicas de Inteligencia Artificial radica en la superación del límite al que habían llegado los positivistas: las inferencias no deben representarse, exclusivamente, recurriendo a la Lógica de Predicados o a inferencias estadísticas simples. La informática ofrece unos formatos de representación que permiten superar las limitaciones semánticas, tanto del lenguaje natural como de los lenguajes formales al uso. Obviamente, esta representación computacional también está afectada por los teoremas de la incompletitud de los sistemas formales: no se busca construir la máquina que acabe con todas las máquinas, sino analizar cómo interpretan los arqueólogos sus evidencias y, lo más importante, por qué lo hacen de ese modo y no de otro.

A buen seguro habrá quien encuentre las páginas anteriores un nuevo ejemplo de palabrería técnica alejada de las preocupaciones de todo buen arqueólogo práctico. Mi único propósito ha sido el de señalar el principal peligro en cualquiera de las ramas de la Ciencia Social: *la trivialización de las inferencias*. Mientras estemos anclados en un modelo verbal del razonamiento y no tengamos conciencia de la diversidad de operaciones asociativas posibles, nos resultará de todo punto imposible comprender la complejidad de los fenómenos sociales.

BIBLIOGRAFÍA

- ALEKSANDER, I., MORTON, H.: 1990 *An introduction to neural computing*. Londres: Chapman and Hall.
- ANDERSON, J.R.: 1983 *The Architecture of Cognition*. Harvard University Press, Cambridge(MA).
- BARCELÓ, J.A.:
- 1992a “Using Intelligent Databases in Archaeology” *Colloque Européen Archéologie et Informatique*. (Saint-Germain-en Laye, France -21-24 Noviembre de 1991). En Prensa.
 - 1992b “El uso de programas de inducción automática en Arqueología de cazadores-recolectores”. *Archi-Diskodon*. No. 2 (Publicación sobre soporte magnético distribuida por Librería Tipo).
 - 1993 “Automatic Problem Solving in Archaeology: a computational framework”. *Archeologia e Calcolatori* No. 4. (en prensa)
- BROWN, D.C., CHANDRASEKARAN, B.: 1989 *Design Problem Solving Knowledge Structures and Control Strategies*. Londres: Pitman.
- BENFER, R., BRENT, E.E., FURBEE, L.: 1991 *Expert Systems*. (Sage University Paper series on Quantitative Applications in the Social Sciences, series n° 07-077). Newbury Park (CA): SAGE Publ.
- BESNARD, P.: 1989, *An introduction to Default Logic*. Nueva York/Berlin: Springer-Verlag.
- BLASIUS, K.H., BURKERT, H.J.: 1989 *Deduction Systems in Artificial Intelligence*. Chichester: Ellis Horwood.
- CARNAP, R.: 1966, *Philosophical foundations of physics*. (trad. castellana, Barcelona: Ediciones Orbis).
- CAUDILL, M., BUTLER, C.: 1990, *Naturally Intelligent Systems*. The MIT Press, Cambridge (MA).
- CLANCEY, W.: 1985 “Heuristic Classification”. *Artificial Intelligence* No. 27: 289-350.
- CLARKE, D.: 1968 *Analytic Archaeology* (Trad. castellana, Barcelona: Editorial Bellaterra).
- CLAXTON, J.B.: 1993 “Geographical Information Systems. The case for adopting parallel architectures”. *Computer Applications and Quantitative Methods in Archaeology 1992*. Compilado por T. Madsen y J. Andresen. Oxford: “Tempus Reparatum” (BAR International Series).
- CHOMSKY, N.: 1965 *Aspects of the Theory of syntax*. Cambridge (MA): The MIT Press.
- DORAN, J.: 1990 *Using Distributed AI to study the emergence of Human Social Organisation*. Research Report. Department of Computer Science, University of Essex.
- FALKENHAINER, B.: 1990 “A Unified Approach to explanation and theory formation”. En J. Shrager y P. Langley (comps.) *Computational models of scientific discovery and theory formation*. Los Altos 5MA): Morgan Kaufmann.
- FELDMAN, J.A., BALLARD, D.H.: 1982 “Connectionist Models and Their Properties”. *Cognitive Science* No. 6: 205-264.
- FRANCFORT, H., LAGRANGE, M.S., RENAUD, M.: 1989 *PALAMEDE. Application des systèmes experts à l'archéologie de civilisations urbaines protohistoriques*. Paris: C.N.R.S.-U.P.R. 315.
- FRIXIONE, M., GAGLIO, S., SPINELLI, G.: 1992 *Representing Concepts in semantic nets: in search of the elusive link*. *Artificial Intelligence*. Chichester: Ellis Horwood.
- GIBBINS, P.: 1990 “BACON bytes back”. En J.E. Tiles, G.T. McKee, G.C. Dean (comps.) *Evolving Knowledge in Natural Science and Artificial Intelligence*. Londres: Pitman.
- GIBSON, P.: 1993 “The potential of hybrid neural network models for archaeofaunal ageing and interpretation”. *Computer Applications and Quantitative Methods in Archaeology 1992*. Compilado por T. Madsen y J. Andresen. Oxford: “Tempus Reparatum” (BAR International Series).
- GILHOOLY, K.J.: 1989 (comp.) *Human and Machine Problem-Solving*. Nueva York: Plenum Press.

- GRÉGOIRE, E.: 1990 *Logiques Non-Monotones et Intelligence Artificielle*. Paris: Hermès.
- HOLLAND, J.H., HOLYOAK, K.J., NISBETT, R.E., THAGARD, P.R.: 1986 *Induction. Processes of Inference, Learning and Discovery*. The MIT Press, Cambridge (MA).
- JOHNSON-LAIRD, P.N.:
 — 1983 *Mental Models*. Cambridge University Press.
 — BYRNE, R.M., 1991 *Deduction*. Hillsdale (NJ): Lawrence Erlbaum.
- JOSEPHSON, J.R., CHANDRASEKARAN, B., SMITH, J.W., TANNER, M.C.: 1987 “A mechanism for forming composite explanatory hypotheses”. *IEEE Transactions on Systems, Man and Cybernetics*, No. 17: 445-454.
- LANGLEY, P., SIMON, H.A., BRADSHAW, G.L., ZYTKOV, J.M.: 1987 *Scientific Discovery. Computational Explorations of the Creative Process*. Cambridge (MA): The MIT Press.
- LAURIÈRE, J.L.: 1986 *Intelligence Artificielle. Resolution de problèmes par l’Homme et la Machine*. Eyrolles, Paris.
- LUKASZEWICZ, W.: 1990 *Non-Monotonic Reasoning. Formalization of Common-sense reasoning*. Chichester: Ellis Horwood.
- MAÍCAS, R.: 1989 “Ejemplos de aplicación de inteligencia artificial en arqueología”. *Cuadernos de Prehistoria y Arqueología de la Universidad Autónoma de Madrid*, No. 16: 73-80.
- MARKEL, J.L.: 1987 *Archaeology and the computer technology revolution*. Ph.D. Dissertation. Buffalo (NY): State University of New York. (Ann Arbor: University Microfilms International).
- MARR, D.: 1982 *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.
- NEWELL, A.: 1990 *Unified Theories of Cognition*. Harvard: Harvard University Press.
- O’RORKE, P. MORRIS, S., SCHULENBURG, D.: 1990 “Theory formation by Abduction: A case study based on the Chemical Revolution”. En J. Shrager y P. Langley (comps.) *Computational Models of Scientific Discovery and Theory Formation*. Los Altos (CA): Morgan Kaufmann.
- OWEN, S.: 1991 *Analogy for Automated Reasoning*. Londres: Academic Press.
- PAO, Y.H.: 1989 *Adaptive pattern recognition and neural networks*. Reading (MA): Addison-Wesley.
- PARTRIDGE, D.: 1990 “Connectionism is better for Engineers than for scientists”. En J.E. Tiles, G.T. McKee y G.C. Dean (comps.) *Evolving Knowledge in Natural Science and Artificial Intelligence*. Londres: Pitman.
- PEARL, J.: 1985 *Heuristics. Intelligent Search Strategies for computer problem solving*. Reading (MA): Addison-Wesley.
- PITRAT, J.: 1990 *Métaconnaissance. Futur de l’Intelligence Artificielle*. Paris: Hermès.
- REYNOLDS, R.G.: 1986 “An adaptive computer model for the evolution of plant collecting and early agriculture in the eastern valley of Oaxaca”. En K. Flannery (comp.) *Guilá Naquitz: archaic foraging and early agriculture in Oaxaca, México*. Nueva York: Academic Press.
- SACERDOTI, (ed): 1977 *A structure for Plans and Behavior*. Nueva York: Elsevier.
- SHASTRI, L.: 1988 *Semantic Networks: An Evidential Formalization and its Connectionist Realization*. Londres: Pitman (Research Notes in Artificial Intelligence).
- SIMON, H.A.: 1972 *Human Problem Solving*. Englewood Cliffs (NJ): Prentice Hall .
- SOWA, J.F.: 1984 *Conceptual Structures: Information Processing in Mind and Machine*. Reading (MA): Addison-Wesley.
 — 1991, (comp.) *Principles of Semantic Networks. Explorations in the Representation of Knowledge*. Los Altos (CA): Morgan Kaufmann.
- STUTT, A., SHENNAN, S.: 1990 “The nature of archaeological arguments”. *Antiquity* vol. 64:766-77.

- THAGARD, P.R.: 1984 "Conceptual combination and scientific discovery" en P. Asquith y P. Kitcher (comps.) *PSA 1984*, (vol. 1). East Lansing, MI: Philosophy of Science Association.
- 1988, *Computational Philosophy of Science*. The MIT Press, Cambridge (MA).
- 1989, "Explanatory Coherence". *Behavioral and Brain Sciences*. No. 12: 435-502.
- 1990, "Modelling Conceptual Change" en *Evolving Knowledge in Natural Science and Artificial Intelligence*.
Compilado por J.E. Tiles, G.T. McKee, G.C. Dean. Londres: Pitman.
- TORASSO, P., CONSOLE, L.: 1988 *Diagnostic Problem Solving. Combining heuristic, approximate and causal reasoning*. London: Chapman and Hall.
- URBACH, P.: 1982 "Francis Bacon as a Precursor to Popper". *The British Journal for the Philosophy of Science*, 33: 113-132.
- 1987, *Francis Bacon's Philosophy of Science*. La Salle (IL): Open Court.
- WATSON, P.J., LEBLANC, S.A., REDMAN, C.L.: 1971 *Explanation in Archaeology. An Explicitly Scientific Approach*. Nueva York: Columbia University Press. (trad. cast. Madrid: Alianza Editorial).
- WEISS, S. KULIKOWSKI, C.: 1990 *Computer systems that learn: Classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. San Mateo (CA): Morgan Kaufmann.
- ZEIDENBERG, M.: 1990 *Neural Networks in Artificial Intelligence*. Chichester: Ellis Horwood.

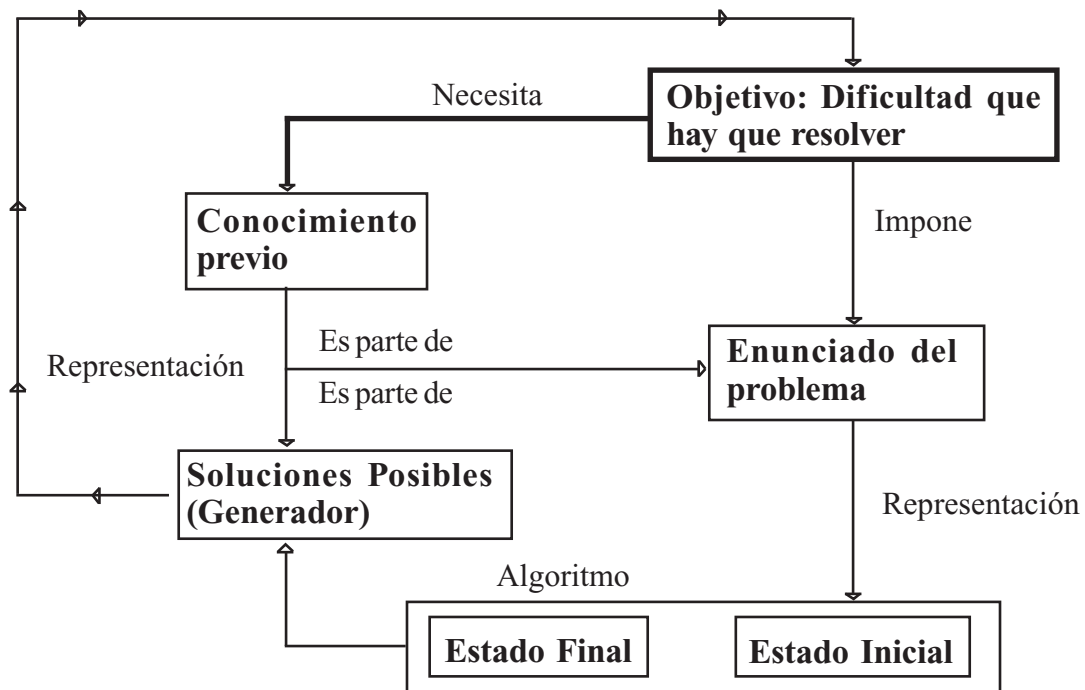


Figura 1. Estructura sistémica y componentes de un problema arqueológico.

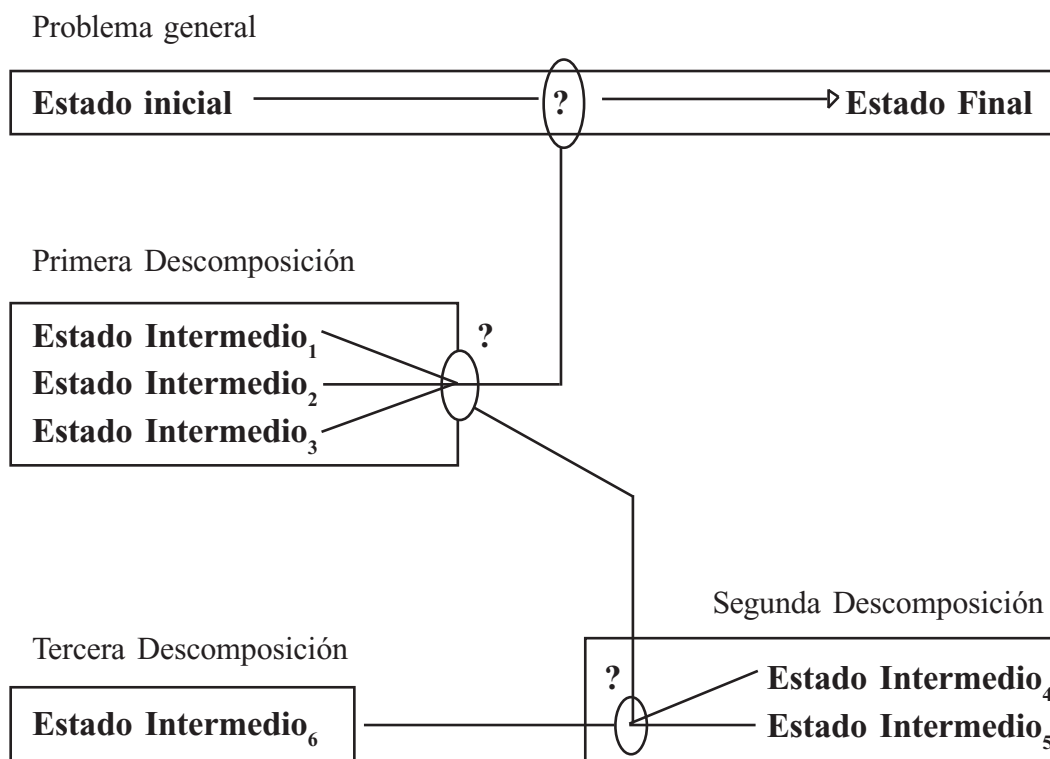


Figura 2. Descomposición jerárquica del proceso de resolución de un problema arqueológico.