

El tratamiento informático de datos funerarios cualitativos: análisis de correspondencias y algoritmo ID3 de Quinlan

Víctor M. Fernández Martínez

Departamento de Prehistoria. Universidad Complutense.
28040 Madrid

Mariano García de la Fuente

Fycsa, Madrid

1. Introducción

Entre los años 1978 y 1981, y totalizando cerca de seis meses de trabajo de campo, la misión arqueológica española en Sudán, dirigida entonces por el Prof. Almagro Basch, llevó a cabo la excavación completa de la necrópolis meroítica antigua de Amir Abdallah, en Abri (Provincia del Norte), pequeña población situada en la margen derecha del río Nilo a unos 200 kms. de la frontera con Egipto.

La necrópolis pertenecía cronológicamente a los primeros siglos meroíticos (III-I a.C.), período muy mal conocido en el momento de los trabajos, en contraste con la información disponible sobre el floreciente período meroítico clásico (I-IV d.C.), del que existían datos (funerarios sobre todo; los asentamientos han sido borrados por la ocupación posterior) muy abundantes en la zona inmediatamente al norte de Abri: la Baja Nubia al sur de Egipto, región al parecer casi totalmente despoblada durante el último milenio antes de nuestra era. Por lo tanto, el registro de las 389 tumbas de la necrópolis, y la comparación posterior de su cultura material con los escasos yacimientos alto-meroíticos publicados desde principios de siglo (siempre inapropiadamente fechados como bajo-meroíticos o simplemente meroíticos), permitió por vez primera la definición histórico-cultural de la zona norte de la Alta Nubia durante ese período (FERNANDEZ, 1984, 1985).

Aparte del estudio de los ajuares funerarios (cerámicas y cuencos de bronce sobre todo), se intentó un análisis global de los aspectos puramente funerarios con los datos, en su mayor parte cualitativos, de la necrópolis. El objeto era determinar la posible correlación entre unas y otras variables, como base de posibles modelos de estructuración de los datos, que a su vez servirían para inferir aspectos sociales y rituales del grupo humano que utilizó la necrópolis. En el momento de llevar a cabo el análisis (año 1982), la falta de un apoyo informático y estadístico apropiado provocó que aquél se limitase a comparaciones bivariantes entre cierto número de variables, aquellas que *a priori* parecían de mayor interés.

Tal carencia es la que hemos tratado de remediar con este trabajo: utilizando un tipo de análisis multivariante muy apropiado para datos cualitativos (A. de Correspondencias), que aunque ya conocíamos en 1982 no pudimos aplicar por no contar con asistencia técnica en el Centro de Cálculo de la UCM, y que hoy ya es posible utilizar directamente sobre un ordenador personal, y comparando sus resultados con los obtenidos mediante un método más reciente,

el sistema generador de reglas ID3 de Quinlan, con el fin de incorporar los actuales procedimientos desarrollados en el campo de la Inteligencia Artificial, y obtener de ello conclusiones generales que sirvan para otros análisis similares que se realicen en el futuro.

2. Los datos: análisis cronológico

Las tumbas de Amir Abdallah eran de pozo alargado, excavado en la gravera de una terraza algo alejada del río, con el cadáver depositado, normalmente en posición extendida, dentro de una cámara, cerrada por muro de adobes, bien al oeste o al este de la fosa. Estos dos tipos eran los más abundantes, pero también existían las tumbas de cámara lateral al norte o al sur de la fosa, y las de nicho simple. El análisis estadístico principal se efectuó con los materiales de ajuar, cerámicas divididas en clases generales que a su vez se clasificaron en tipos de acuerdo con su forma y decoración: cerámica a mano, incisa e impresa (24 tipos); cerámica a mano grosera (6 tipos); cerámica a mano con engobe rojo (4 tipos); cerámica a torno con engobe bruñido (18 tipos); y cerámica helenística, siendo la importada de Aswan la más importante. Los cuencos de bronce se dividieron en lotiformes, hemiesféricos, exvasados y parabólicos.

Como era lógico, el primer estudio se centró en el contexto espacial de los diferentes tipos dentro de la necrópolis. Partiendo de la base de que cada tipo tuvo una posición cronológica concreta, es decir, se fabricó sólo durante un cierto tiempo, era de esperar que su aparición en las tumbas fuera igualmente concentrada. Debido a que el espacio disponible para las tumbas era ilimitado (la necrópolis, al contrario que en otros casos, se había erigido en pleno desierto y no en la llanura aluvial), éstas seguían una disposición espacial ordenada en filas, sin existir ninguna superposición: es decir, las primeras habían sido excavadas al norte y las últimas al sur. No obstante, aunque el orden dentro de cada fila de tumbas era el «correcto», una detrás de otra, entre las diferentes filas (26 en total, tal vez correspondientes a distintas unidades de parentesco) existían corrimientos temporales, es decir, unas iban más adelantadas que otras. Para corregir estas diferencias, se siguió el sistema (FERNANDEZ, 1985: 156-65) de compaginar la posición de los tipos en las filas, deslizando unas respecto a otras en un diagrama hasta que los bronceos y vasijas de un mismo tipo quedaran lo más cerca posible unos de otros (el principio «de concentra-

ción»; KENDALL, 1971: 217, siguiendo los trabajos de seriación de tumbas predinásticas de PETRIE, 1899).

Al final obtuvimos un diagrama con las filas de tumbas colocadas tal como debieran haber estado si se hubiera seguido la misma «velocidad de deposición» en todas ellas. De esta manera, y en teoría, las tumbas colocadas a la misma altura debieron ser excavadas en el mismo momento, aunque pertenecan a distintas filas. Por lo tanto, ahora era posible cortar el diagrama en «rodajas» perpendiculares a las filas, cogiendo en cada una tumbas de distintas alineaciones, para formar grupos aproximadamente contemporáneos, que llamamos «generaciones». Así, y exclusivamente como un recurso analítico con fines heurísticos, la necrópolis quedó dividida en once generaciones que, para rizar el rizo, fueron supuestas de 25 años cada una y colocadas entre las fechas límite de 300 y 25 a.C.

Posteriores estudios de seriación permitieron comprobar la bondad relativa de tal sistema: utilizando como unidades las generaciones, un análisis de correspondencias y otro de proximidades (MDSCAL) mostraron a los grupos relativamente bien ordenados de principio a fin (con algunas inversiones, como la generación 1, que aparece desplazada hacia el centro de la serie), y la matriz de datos «diagonalizada», es decir con los tipos concentrados en la diagonal como es norma cuando existe un orden cronológico (FERNANDEZ, 1989).

3. Los datos cualitativos: análisis bivalente

Aunque durante la excavación se realizó una descripción textual de cada tumba, empleando hojas-formulario especialmente preparadas para ello, en la publicación definitiva preferimos la descripción formal, mediante códigos alfanuméricos en cada variable. De esta manera, y en cuanto al ahorro de espacio, se pasó de cerca de 800 folios a únicamente 26 páginas (FERNANDEZ, 1985: 742-767). Por otro lado, y más importante que la ventaja anterior, la base de datos estaba ya preparada para un tratamiento estadístico susceptible de ser llevado a cabo manualmente o con un ordenador.

Cada tumba era definida mediante 25 variables: número, generación a la que pertenece, sexo y edad del difunto, estado de la tumba (intacta o violada), tipo general (visto antes), tipo de fosa, de suelo en fosa y cámara, de cerramiento de la cámara, presen-

cia o ausencia de superestructura sobre la tumba (restos de pirámide de adobes), dimensión de la tumba (longitud total, variable cuantitativa que fue «discretizada» en decímetros), posición del cadáver, posición del cráneo, manos y piernas, tipo de acomodamiento del cadáver (varias clases de ataúd o lecho), presencia/ausencia de cuentas de collar, faldellín de cuero y sudario de tela, posición del ajuar dentro de la cámara, tipo cultural del ajuar y número de unidades que componen el mismo, tipo de violación (en cuanto a intensidad) sufrida por la tumba cuando era el caso, así como la forma de destrucción del muro, y finalmente los tipos de fragmentos cerámicos encontrados en el relleno.

Para cada una de estas variables existían lógicamente varios estados, que se codificaron mediante números distintos (podrían haber sido letras). Así, por ejemplo, en la posición del cadáver, «1» significa extendido boca arriba, el tipo más frecuente, «2» con las piernas flexionadas hacia el norte, «3» flexionadas hacia el sur, etc. En general se reservó el número «0» para los casos en que no existía información respecto a esa variable: por ejemplo, cuando una tumba había sido violada intensamente, removiendo el cadáver, no se sabe cuál fue la posición original del mismo.

Como ya dijimos, el camino analítico seguido con estos datos fue la comparación bivariada entre ellos tomados de dos en dos. Esto, que fue hecho manualmente, supuso como es lógico un enorme trabajo, que además fue incompleto, puesto que sólo se realizaron unas treinta comparaciones de las 276 posibles (24x23/2 combinaciones). Con todo, el análisis no fue hecho ciegamente, sino que se contrastaron aquellas variables sobre las que en principio existían indicios de correlación, por lo que los resultados fueron relativamente interesantes.

Algunos hechos ya observados durante la excavación fueron comprobados en este análisis: correlación entre cronología (generaciones) y tipos de tumba, al principio exclusivamente de cámara al oeste y luego de los restantes tipos; correlación entre tipos de tumba y violación de las mismas, puesto que los ladrones de ajuares funerarios (cuentas de collar) habían actuado mucho más en las tumbas de cámara al oeste que en las que tenían la cámara al este (con lo cual se relacionaba también la cronología: más violación al principio que al final).

Otros resultados fueron: 1) presencia relativamente mayor de adultos en las primeras generaciones, y de maduros en las últimas; 2) no relación entre sexo y tipo de tumba, pero más seniles en las cámaras al oeste, más maduros en cámaras al este y más juveni-

les en cámara lateral y nicho; 3) cámaras al oeste en relación con fosas en rampa o escalones-rampa, y cámaras al este en relación con fosas de pared vertical; 4) suelos planos en cámaras al este, suelos pendientes en cámaras al oeste; 5) muros de cerramiento más elaborados (doble o triple hilada de adobes) en cámaras al oeste, y menos (hilada simple, amontonamiento irregular de adobes o piedras) en cámaras al este; 6) relación entre la presencia de superestructura y la riqueza del ajuar; 7) disminución del tamaño de las tumbas con el paso del tiempo; 8) posiciones «arcaicas» del cadáver (flexionado, contraído) sólo al principio de la necrópolis; 9) cráneo hacia arriba más frecuente en varones y piernas cruzadas algo más abundante en mujeres; 10) acomodaciones «arcaicas» (lecho de madera) más frecuentes en las primeras generaciones y en tumbas femeninas; 11) relación entre presencia de ataúd y riqueza del ajuar; 12) envoltura de cuero en pélvis sólo en últimas generaciones y sólo en varones; 13) mayor riqueza de ajuar en varones, sobre todo maduros y seniles; 14) cuentas de collar en niños y algo más en mujeres que en hombres (FERNANDEZ, 1985: 182-297; FERNANDEZ, 1984: 51-57).

En todas esas correlaciones se construyeron tablas de contingencia con las frecuencias de aparición conjunta de los estados de cada dos variables. Aunque el contraste del chi-cuadrado, calculado siempre, sólo en muy pocas ocasiones fue significativo (indicando una clara asociación o disociación), en general se tuvieron en cuenta las «tendencias» hacia la correlación de variables para las conclusiones del trabajo. En ellas se deducía un modelo general de organización del ritual funerario de la necrópolis, nunca realizado antes con ese detalle en el registro meroítico, con el fin de compararlo en el futuro con otras investigaciones para llegar a una mejor comprensión de ese aspecto de la historia nubia.

4. Análisis de correspondencias

Esta clase de análisis multivariante, que resulta muy apropiado para datos cualitativos o tablas de contingencia, fue ideado, en la forma que utilizamos aquí, por Jean Paul Benzecri a comienzos de los años sesenta (BENZECRI, 1973), aunque existían sistemas parecidos en el ambiente anglosajón desde la década de los treinta (GREENACRE, 1981). Se trata de un método de «reducción de datos» desde un espacio multidimensional a otro de dimensión menor (normalmente de dos ejes), con unos principios teó-

ricos similares a los del Análisis de Componentes Principales (ACP), pero con la particularidad de que la distancia entre casos y variables no utiliza el coeficiente «r» de Pearson o la covarianza, sino la métrica del chi-cuadrado. Otro de sus aspectos, y tal vez su principal ventaja sobre otros métodos, es que la representación incluye en el mismo diagrama las variables y los casos, de forma que no sólo se aprecia la correlación que existe entre las primeras, sino también la que éstas tienen con los casos (de ahí el nombre de «correspondencias»). En la actualidad es uno de los métodos de utilización más frecuente en arqueología y otras ciencias (BØLVIKEN et al., 1982; RINGROSE, 1988).

El programa utilizado fue el AFC, desarrollado en Basic para un micro-ordenador TRS-80 por Foucart (1982) y adaptado para un PC compatible por el primer firmante de este trabajo. Debido a sus características, los datos cualitativos codificados tuvieron que convertirse previamente en dicotómicos (mediante un programa especial en Basic): es decir, cada estado de variable se convirtió en una variable en sí misma, con un estado de presencia y otro de ausencia, indicados por «1» y «0». Por ejemplo, la variable «tipo de tumba», con los cinco estados que antes vimos, se convierte en cinco variables dicotómicas, de forma que si una tumba es del tipo «2» (cámara al este), se representará por «01000», si es del tipo «3» por «00100», etc.

Para el análisis se decidió prescindir de las tumbas violadas, puesto que su información es muy fragmentaria. De las 243 tumbas restantes, decidimos también excluir las tumbas infantiles, porque un análisis de correspondencia con todas ellas indicó que la mayor variabilidad correspondía a la diferencia entre tumbas de adulto y de niño, oscureciéndose los modelos presentes en el conjunto de tumbas grandes. Con ello la muestra de análisis quedó reducida a 206 tumbas. También fueron retiradas algunas variables, como la colocación del ajuar (prácticamente una constante, casi siempre a la entrada de la cámara), el tipo cultural del mismo (ya tenido en cuenta para la separación en generaciones), o los fragmentos cerámicos del relleno.

El resultado (figs. 1 y 2) es interesante por varias razones, a pesar de que la reducción es poco satisfactoria en cuanto a la «inercia» (concepto equivalente a la varianza en el ACP), de la cual sólo se concentra un 14 % en los dos primeros ejes (con todo, las variables tomadas de dos en dos sólo tendrían un 3 % de inercia por término medio). En ambas figuras se aprecia una forma semicircular o de «herradura» en

la colocación, tanto de las tumbas como de las variables (el gráfico conjunto se ha dividido en dos por claridad). Esto ya sugiere en principio una variación cronológica, que se confirma al examinar la colocación de la generación de cada una (fig. 2): se aprecia un orden aproximado desde la derecha hasta la izquierda, con algunas pequeñas inversiones.

Lo mismo se obtiene del análisis de las variables más representadas en cada eje; el eje o factor 1 tiene, en el lado positivo, los siguientes estados de variable: fase 1 (primeras generaciones), tumbas de cámara al oeste (CW), acceso en rampa (RAM, ERA), mayor tamaño de tumbas (60, 70 y 80 dms.), posición del cadáver «despatarrado» (DES), ataúd antropomorfo (ANT) o lecho de madera (LEC), todos ellos aspectos antiguos en la necrópolis. El mismo eje, en sentido negativo, «carga» por el contrario en estados de variable modernos: fase 3 (últimas generaciones), tumbas de cámara al este, lateral y de nicho (CE, CL, NI), acceso vertical (VER), edad madura (MAD), tumbas pequeñas (30 y 40 dms.), mayor cantidad de ajuar (2 y 3 ó más unidades), etc.

El segundo eje, vertical, es de interpretación algo más difícil: en sentido positivo mezcla caracteres antiguos (fase 1, ataúdes arcaicos) y modernos (nichos, suelo plano, ausencia de muro, tumbas pequeñas) y en sentido negativo caracteres modernos y otros indiferentes para esta distinción. Ahora bien, arriba están la edad juvenil (JUV), ausencia de ajuar (0) y abajo la edad madura y el mayor ajuar, por lo que deducimos que su significado está en relación con la riqueza de las tumbas, arriba las pobres (incluyendo las antiguas, que son todas pobres) y abajo las ricas. Por lo tanto, en la reducción de datos la variabilidad en ese aspecto interfiere con la cronológica, más importante, dando como resultado la forma curva del gráfico.

De todo lo anterior se puede deducir la división de la necrópolis, tanto de las tumbas como de los estados de variable, en tres o cuatro grupos diferentes (fig. 1): 1) grupo «antiguo», colocado arriba a la derecha, definido por las acomodaciones y la posición del cadáver arcaicas, cámara al oeste, sin ajuar, tamaño grande, fosa en rampa o similar, suelo pendiente, muro de hiladas dobles y triples y edad predominante de adulto; 2) grupo «moderno rico», colocado abajo en el centro, definido por cámara al este, tamaño medio (40-50 dms.), acceso vertical, máximo ajuar, superestructuras, muro irregular, cadáver con piernas cruzadas (CRU) y edad predominante de maduro; 3) grupo «moderno pobre», arriba a la izquierda, definido por tumbas de cámara lateral y de nicho, pared vertical, muros simples o sin muro, tamaño

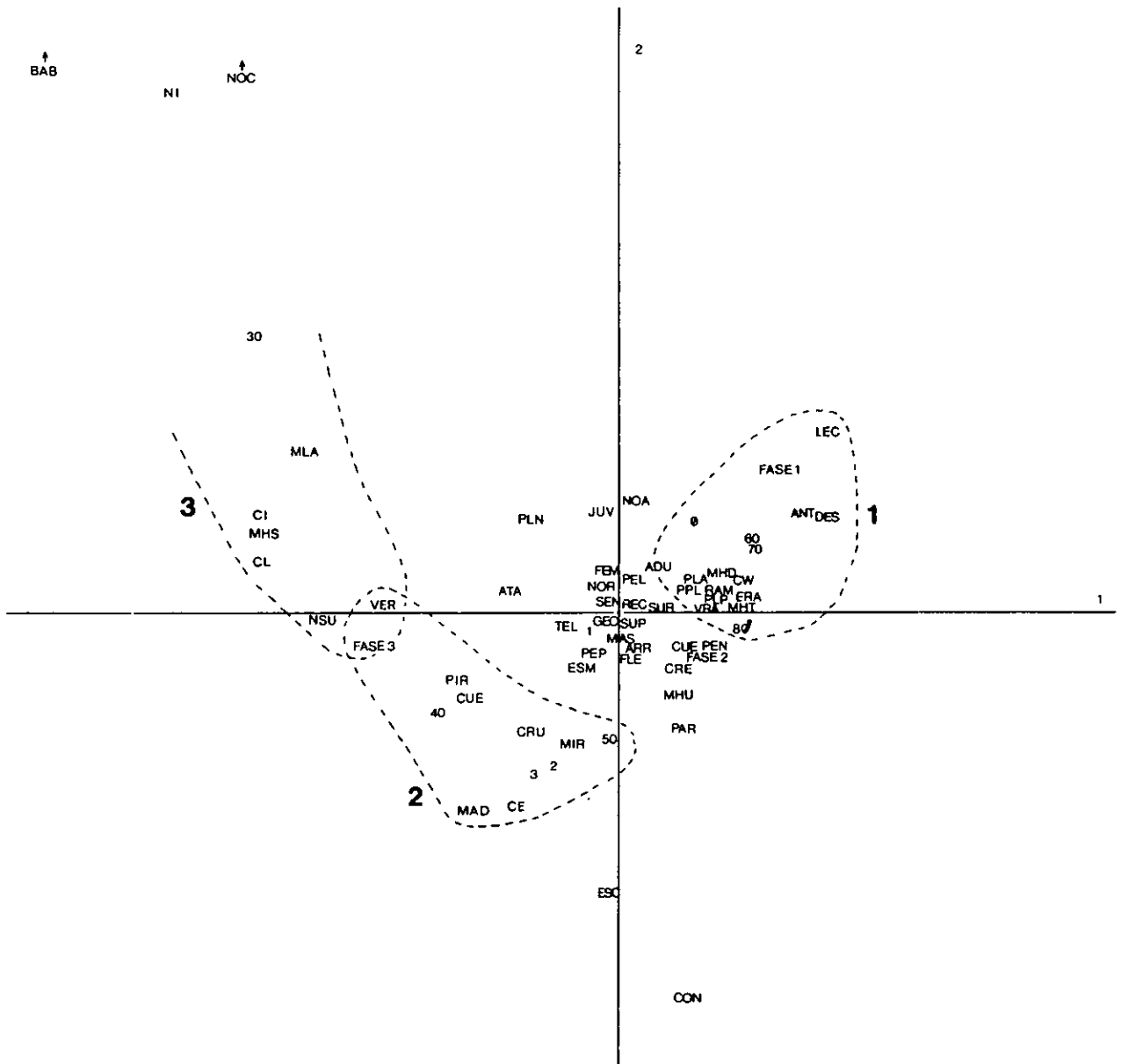


Fig. 1.—Resultados en los dos primeros ejes del análisis de correspondencia de 206 tumbas meróticas definidas por 67 estados de variables cualitativas. En este diagrama se han representado únicamente los estados de variable, según los siguientes códigos. FASE 1: generaciones 1-3; FASE 2: generaciones 4-8; FASE 3: generaciones 9-11. MAS: sexo masculino; FEM: sexo femenino. JUV: edad juvenil; ADU: adulto; MAD: maduro; SEN: senil. CW: cámara al oeste; CE: al este; CL: lateral; NI: nicho. VER: acceso vertical; RAM: rampa; VRA: vertical y rampa; ESC: escalones; ERA: escalones y rampa. PLN: suelo plano; PEN: pendiente; PPL: pendiente/plano; PLP: plano/pendiente; ESM: escalón bajo el muro. MHU: muro de hilada única; MHD: hilada doble; MHT: triple; MIR: irregular; MHS: hilada simple lateral; MCI: cista; MLA: lajas; NOC: sin cierre. PIR: superestructura (pirámide). 20: dimensión menor de 2 m.; 30: 2-3 m.; 40: 3-4 m.; 50: 4-5 m.; 60: 5-6 m.; 70: 6-7 m.; 80: 7-8 m. SUP: cadáver en posición supina; FLE: flexionado; CON: contraído; DES: despatarrado; NSU: según norte-sur; BAB: boca abajo; CRE: cráneo al este. ARR: cráneo hacia arriba; NOR: hacia el norte; SUR: al sur. PEL: manos sobre pelvis; PAR: paralelas; PEP: alternas. REC: piernas rectas; CRU: cruzadas. ATA: forma del ataúd sin determinar; GEO: geométrico; ANT: antropomorfo; LEC: lecho de madera; PLA: plancha; NOA: sin ataúd. CUE: sudario de cuero. TEL: sudario de tela. CUN: cuentas de collar. O: sin ajuar; 1: una unidad de ajuar; 2: dos unidades; 3: tres o más unidades.

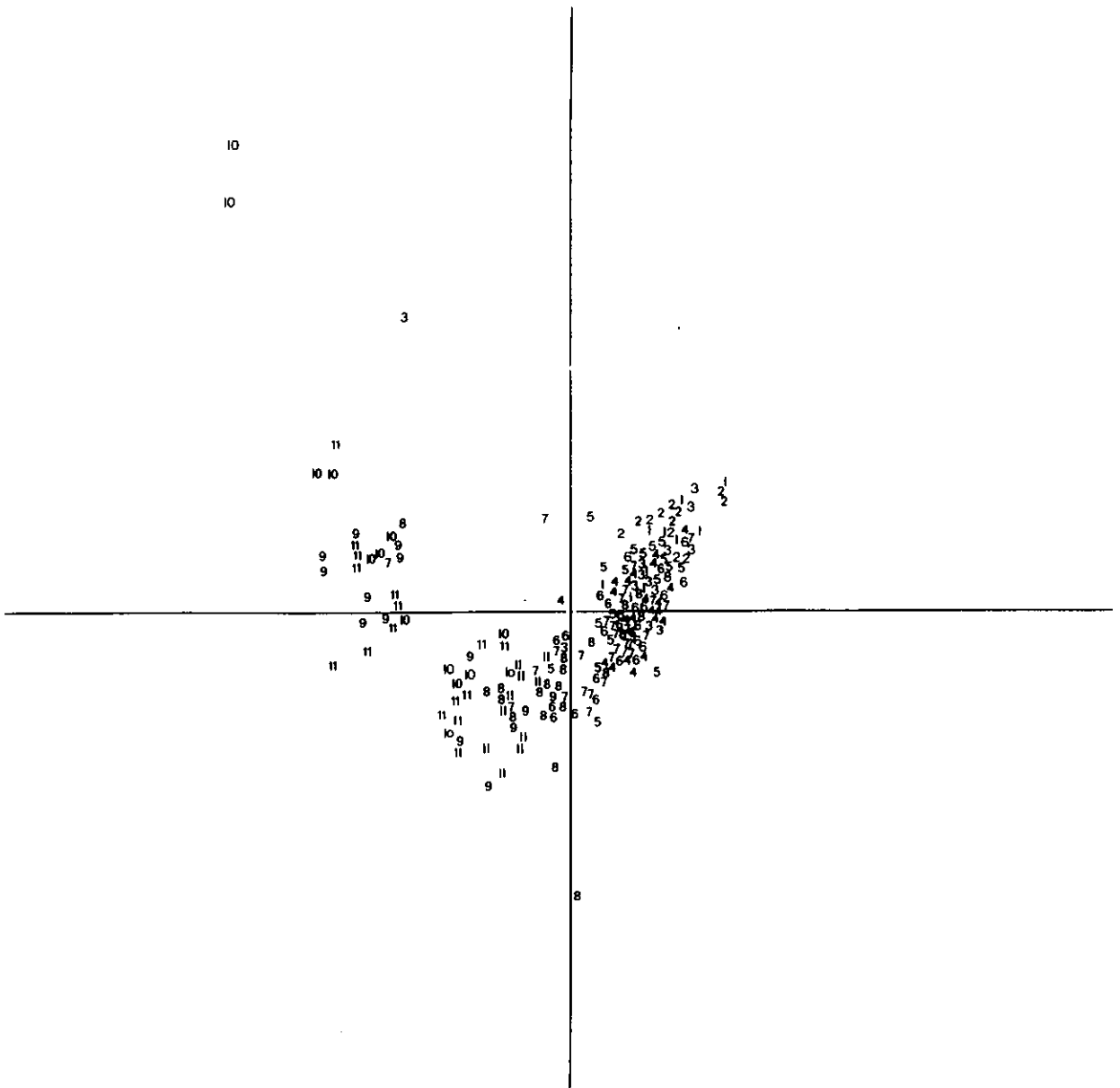


Fig. 2.—Resultado en los dos primeros ejes del análisis de correspondencias de 206 tumbas meroíticas definidas por 67 estados de variables cualitativas. En este diagrama se representan las tumbas de la necrópolis, indicando en cada caso la generación cronológica a la que pertenece (generaciones de 1 a 11, ca. 300-25 a. de C.).

pequeño (30 dms.), poco ajuar en comparación con el grupo anterior (pero tiene más que el primer grupo), y algunas posiciones «raras» de cadáver (boca abajo, BAB, orientado según norte-sur, NSU), que en alguno de los casos es probable que correspondiera a un sacrificio humano.

Entre estos tres grupos existe una aglomeración de tumbas y variables justo en el centro, alrededor de la confluencia de los dos ejes, que están ahí precisa-

mente porque representan el «perfil» de las variables, su valor medio, y por ello pueden aparecer en cualquiera de los otros grupos. Este grupo «típico» o «central» está definido por la fase 2 (generaciones centrales), ningún tipo específico de tumba (luego puede tener todos), suelo con escalón bajo el muro (ESM, el más abundante), muro de hilada única (MHU), posición supina (SUP), cráneo y manos en cualquier posición, piernas paralelas, ataúd y no

ataud, ambos sexos (variable de relativa poca importancia diferenciadora), ajuar medio, edad juvenil y senil, sudario de tela, etc. Tales variables, y las tumbas de este grupo, se podrían haber eliminado en un análisis subsiguiente si se hubiese considerado necesaria una mayor claridad en la representación gráfica. Como se ha visto, son las variables no significativas en cuanto al doble modelo (cronológico-social) descubierto en la necrópolis.

En los grupos anteriores no existe una presencia al cien por cien de todos los estados de variable, pero sí una cierta «correspondencia»: por ejemplo, y como caso más extremo, las piernas cruzadas sólo representan el 13 % del segundo grupo (3 casos de 22), pero esas tumbas son el 50 % de todas las que tienen el cadáver así colocado en toda la necrópolis. Es decir, el método no sólo funciona cuando existe una correspondencia total (si todos los casos de un grupo son exactamente iguales) sino que también sirve para indicar tendencias o grupos más o menos perfectos, incluso en datos con una estructura «débil» como son los de Amir Abdallah, necrópolis de un gran conservadurismo cultural y ritual, típico de la zona.

En conclusión, el análisis de correspondencias ha permitido inferir un cierto modelo en los datos funerarios, que ya se había intuído en parte previamente mediante el análisis cronológico y bivariado (al principio tumbas pobres pero iguales; luego aparecen las diferencias, probables reflejos de una estratificación social incipiente motivada por el mayor contacto con el mundo helenístico del norte, mediante el comercio, y con el meroítico del sur, que fue paulatinamente incrementando el control político sobre la zona), pero que ahora se ha llevado a cabo con mucho menos trabajo, completándose con la inclusión de otras variables antes no apreciadas, y se ha podido «comprimir» en un sólo gráfico para su exposición pedagógica.

5. Inducción automática: ID3

En la construcción de Sistemas Expertos, uno de los ya clásicos «cuellos de botella» consiste en adquirir las «reglas» que permiten a un experto solucionar los problemas de su dominio. Es por ello que distintos autores han tratado de confeccionar mecanismos que permitan inducir dichas reglas a partir de tablas cualitativas en las que se recogen los distintos casos que ya ha solucionado el experto.

Las «reglas» que se obtienen mediante un mecanismo de inducción automática van a tener presumiblemente capacidad predictiva para resolver

con la misma eficacia que el experto los nuevos casos que se formulen. Al mismo tiempo cada regla nos está mostrando las relaciones entre grupos de estados de variables.

En nuestro caso, el aplicar un mecanismo de inducción sobre los datos de la necrópolis no está encaminado a obtener reglas con capacidad predictiva, sino a encontrar relaciones entre estados de variables.

La diferencia principal de este mecanismo de inducción con el método de análisis de correspondencias, como veremos, radica en que las agrupaciones entre estados de variables no son «espontáneas», sino que están enfocadas a la discriminación de los estados de una de las variables consideradas.

El mecanismo de inducción que hemos empleado es una implementación en LISP del algoritmo ID3 de Quinlan (QUINLAN, 1983; 1986), que «corre» sobre un ordenador personal compatible. En fecha más reciente, una versión avanzada de dicho algoritmo ha ofrecido resultados prometedores en el campo del diagnóstico médico (QUINLAN, 1987), concretamente en el diagnóstico de enfermedades hormonales.

Para que dicho mecanismo pueda aplicarse es necesario tomar a una de las variables como «diagnóstico», o variable de la cual interesa discriminar sus distintos posibles estados. A partir de esto lo que se intenta es ver qué combinaciones de estados del resto de las variables discriminan los posibles estados de la variable «diagnóstico».

Para generar las combinaciones de estados de las variables, éstas se van seleccionando en cada momento en función de la «entropía» que resulta de realizar una partición sobre el conjunto de datos mediante los distintos estados de cada variable. Así cada variable ante un conjunto de datos generará un entropía determinada. ID3 selecciona entonces aquella variable que genera una menor entropía.

En el estudio de la necrópolis hemos tomado sucesivamente como variable a discriminar cada una de las que se consideran en la tabla cualitativa descrita anteriormente, y hemos hecho que el algoritmo agrupara los estados del resto de las variables, de forma que se discriminaran unívocamente en función de dichas agrupaciones («reglas») los estados de la variable que en ese momento se tomaba como diagnóstico.

Los resultados ciertamente no han sido todo lo brillantes que hubiésemos deseado. Las posibles razones de este hecho son las siguientes:

i) El algoritmo ID3 es mejorable. El propio Quinlan ha mejorado dicho algoritmo con el empleo del C4 (QUINLAN, 1987).

ii) La necrópolis parece corresponder a una sociedad poco estratificada, según se desprende de la escasa variabilidad de los datos funerarios. Esto hace que sean más difíciles de discriminar los distintos estados de las variables, aunque posiblemente en estos casos es cuando más se necesitan mecanismos automáticos de agrupación de los mismos.

iii) Los datos pueden contener «ruido», es decir, variaciones aleatorias no relacionadas con el sistema social y religioso que tratamos de identificar, al cual oscurecen. Este ruido hace que el algoritmo genere un número grande de agrupaciones, pero con menor potencia de clasificación.

No obstante, el análisis llevado a cabo por el mecanismo de inducción permite llegar a ciertas conclusiones, algunas de las cuales ya se habían intuido en el primer momento del análisis de datos, como pueden ser: 1) primera fase con tipo de tumba de cámara al oeste; 2) presencia de cuero en individuos de sexo masculino; 3) relación entre la cantidad de ajuar y las generaciones (más ajuar en la última fase); 4) presencia de cuentas en individuos infantiles o de sexo femenino; 5) ausencia de superestructura en fosas de cámara al oeste sin ajuar (correspondientes a tumbas antiguas); 6) acomodación en lecho preferentemente en mujeres, etc.

Una ventaja de este método frente al análisis de correspondencias es que no es necesario convertir las variables originales en bivalentes, sino que se puede aplicar el método sobre la estructura de datos original.

Al igual que en el análisis de correspondencias, se decidió prescindir de las tumbas violadas. Adicionalmente cuando se seleccionaba una variable como discriminante, se eliminaban también los casos en los que no se conocía el estado de dicha variable. Esto es, si se pretendía discriminar la edad del individuo, se eliminaban aquellas tumbas donde por alguna razón no se había logrado conocer el estado de dicha variable.

Por lo general la aplicación de ID3 ha generado demasiadas agrupaciones de estados de variables, o reglas discriminadoras, y al mismo tiempo con poca potencia de discriminación (se entiende por potencia de discriminación de una regla el porcentaje de casos que son clasificados por la misma).

En el ejemplo (a) se aprecia (ver más abajo), si observamos las tres últimas reglas, que ID3 ha necesitado recurrir a la variable que indica la posición de las MANOS para poder «aislar» un solo caso de FASE 2, con lo que se obtienen dos reglas para FASE 1 con menor potencia clasificatoria. Más claro es el ejemplo (b) donde por un solo caso de FASE 2, ID3 recurre a la variable EDAD generando cinco reglas para FASE

3, evidentemente disminuyendo la potencia de clasificación de las mismas. Aun así queda patente que la agrupación de estados de variables (TIPO C_LATERAL_NS) (ATAUD GEO) discrimina relativamente bien el estado de la variable FASE.

Aunque en los datos a analizar no se eliminaron originariamente los individuos infantiles, un análisis posterior de algunas variables eliminando dichos datos realizaba agrupaciones de estados de variables, en algunos casos, notablemente diferentes, lo que confirma que este número reducido de tumbas generaba un cierto factor de distorsión cuando se precisaba la variable SEXO para discriminar alguna otra variable.

En resumen, podemos decir que el método de análisis empleando ID3 no ha proporcionado, salvo excepciones, unas agrupaciones de variables con la potencia de discriminación que esperábamos. No obstante los resultados han permitido constatar algunas de las conclusiones obtenidas con otros sistemas, mediante un examen más automatizado de los datos, y por lo tanto más libre de error.

Para terminar, exponemos a continuación algunos fragmentos de agrupaciones realizadas por ID3 tomando como variable a discriminar la FASE.

LL(a)
 ((TIPO C_OESTE) (ATAUD LECHO) (EDAD NO))

FASE 1 2/52

((TIPO C_OESTE) (ATAUD LECHO) (EDAD JUV))

FASE 2 2/121

((TIPO C_OESTE) (ATAUD LECHO) (EDAD ADU) (MANOS PEL))

FASE 1 6/52

((TIPO C_OESTE) (ATAUD LECHO) (EDAD ADU) (MANOS PAR))

FASE 1 1/52

((TIPO C_OESTE) (ATAUD LECHO) (EDAD ADU) (MANOS RR))

FASE 2 1/121

.

.

.

LL(b)
 ((TIPO C_LATERAL_NS) (ATAUD GEO) (EDAD ÑO))

FASE 2 1/121

((TIPO C_LATERAL_NS) (ATAUD GEO) (EDAD ÑF))

FASE 3 3/71

((TIPO C_LATERAL_NS) (ATAUD GEO)

(EDAD JUV))
 FASE 3 1/71
 ((TIPO C_LATERAL_NS) (ATAUD GEO)
 (EDAD ADU))
 FASE 3 8/71
 ((TIPO C_LATERAL_NS) (ATAUD GEO)
 (EDAD MAD))
 FASE 3 2/71
 ((TIPO C_LATERAL_NS) (ATAUD GEO)
 (EDAD SEN))
 FASE 3 4/71

- FOUCART, T.
 1982 *Analyse Factorielle: programmation sur micro-ordinateurs*. Masson, París.
- GREENACRE, M.J.
 1981 «Practical Correspondence Analysis», en BARNETT, V. (ed.), *Interpreting multivariate data*, John Wiley and sons, Chichester, pp. 119-146.
- KENDALL, D.G.
 1971 «Seriation from abundance matrices», en HODSON, F.R.; KENDALL, D.G.; TAUTU, P. (eds.), *Mathematics in the archaeological and historical sciences*, Edinburgh U.P., pp. 215-252.
- PETRIE, W.M.F.
 1899 «Sequences in prehistoric remains», *Journal of the Anthropological Institute of Great Britain and Ireland*, 29, 295-301.
- QUINLAN, J.R.
 1983 «Learning efficient classification procedures and their application to chess endgames», en MICHALSKI, R.S.; CARBONELL, J.G.; MITCHELL, T.M. (eds.), *Machine learning: an I.A. approach* (vol. I), Tioga, Palo Alto, pp. 463-482.
 1986 «The effect of noise on concept learning», en MICHALSKI, R.S.; CARBONELL, J.G.; MITCHELL, T.M. (eds.), *Machine learning: an I.A. approach* (vol. II), Morgan Kaufman, Los Altos, pp. 146-166.
 1987 «Inductive Knowledge Acquisition: a case study», en QUINLAN, J.R. (ed.), *Applications of Expert Systems*. Addison-Wesley.
- RINGROSE, T.
 1988 «Correspondence analysis as an exploratory technique for stratigraphic abundance data», *Computer and Quantitative Methods in Archaeology 1987*, BAR international Series 393, pp. 3-14.

BIBLIOGRAFIA

- BENZECRI, J.-P.
 1973 *L'analyse de correspondances* (volumen 2.º de *L'analyse des données*). Dunod, París.
- BØLVIKEN, E.; HELSKOG, E.; HELSKOG, K.; HOLM-OLSEN, I.M.; SOLHEIM, L.; BERTELSEN, R.
 1982 «Correspondence analysis: an alternative to principal components», *World Archaeology*, 14(1), 41-60.
- FERNANDEZ MARTINEZ, V.M.
 1984 «Early Meroitic in Northern Sudan: the assessment of a Nubian archaeological culture», *Aula Orientalis*, II(1), 43-84.
 1985 *La cultura alto-meroitica del Norte de Nubia*. Tesis doctoral 132/85. Universidad Complutense, Madrid.
 1989 «Searching for structure within a Meroitic cemetery», en DONADONI, S.; WENIG, S. (eds.), *Studia Meroitica 1984*, Berlín, 469-480.

