

FACULDADE DE LETRAS
INSTITUTO DE ARQUEOLOGIA

CONIMBRIGA

VOLUME XVI



UNIVERSIDADE DE COIMBRA

1977

CARACTERIZAÇÃO DE CERÂMICAS ARQUEOLÓGICAS MEDIANTE ANÁLISE POR ACTIVAÇÃO COM NEUTRÕES TÉRMICOS. CLASSIFICAÇÃO DAS CERÂMICAS POR MÉTODOS DE TAXONOMIA NUMÉRICA

RESUMO

Discutem-se alguns aspectos da aplicação do método instrumental de activação com neutrões térmicos na análise de cerâmicas arqueológicas, em particular do método que está a ser utilizado no Laboratório de Física e Engenharia Nucleares no âmbito dum projecto de investigação recentemente iniciado de colaboração com alguns Museus e Centros de Arqueologia portugueses. Discutem-se, além disso, alguns métodos de taxonomia numérica, utilizados para classificar as cerâmicas com base nos resultados obtidos nas análises radioquímicas, nomeadamente alguns métodos de análise de grupos, hierárquicos e não-hierárquicos, e um método de ordenação.

1. INTRODUÇÃO

Um dos acontecimentos científicos que maior influência exerceram para que arqueólogos, mesmo os mais conservadores, reconhecessem a importância da física e da química no desenvolvimento da sua própria disciplina foi, sem dúvida, a descoberta da existência de ^{14}C (carbono 14) na natureza, feita por Libby¹, em 1946, e da possibilidade da sua aplicação para datar materiais.

Por outro lado, foram os estudos originais de seriação formulados matematicamente em 1951 por Brainerd², um arqueólogo, de colaboração com Robinson³, um matemático, bem como o próprio método de datação do ^{14}C , particularmente em virtude

da necessidade de se interpretarem os resultados que foram sendo obtidos nas suas sucessivas aplicações, o que exige o recurso a técnicas de estatística, que demonstraram inequivocamente a importância da matemática no mencionado desenvolvimento.

A importância desta última disciplina veio a acentuar-se ainda mais a partir do fim da década de 50, com o advento dos computadores, visto que a análise de dados exige, por via de regra, não só numerosíssimas repetições de cálculos, embora simples, mas também a realização dalguns cálculos complexos que, sem o auxílio dessas máquinas e de técnicas adequadas de informática, se tornaria difícil ou mesmo impraticável realizar.

Desde então, o número de contribuições para o desenvolvimento da arqueologia dadas por físicos, químicos, matemáticos e investigadores doutras disciplinas, quer incentivados por arqueólogos, quer por iniciativa própria, tem aumentado consideravelmente de ano para ano, sendo de prever que continue a aumentar cada vez mais no futuro. Disso se poderá tomar consciência folheando revistas, actas de simpósios e livros sobre essa matéria, como por exemplo os citados na bibliografia⁴⁻¹⁰.

Neste artigo ocupar-nos-emos apenas dalguns aspectos da aplicação do método de análise por activação em arqueologia, o qual tem servido de base a algumas dessas contribuições. Vários exemplos da sua aplicação neste domínio constituíram já motivo de revisão numa revista portuguesa de arqueologia¹¹.

Focaremos particularmente a nossa atenção no método instrumental por activação com neutrões térmicos e na sua aplicação à análise de artefactos de cerâmica, uma vez que foi recorrendo àquele método instrumental aplicado à caracterização desses artefactos, que se iniciou muito recentemente um projecto de investigação em arqueologia, no Laboratório de Física e Engenharia Nucleares (LFEN), de colaboração com alguns arqueólogos.

Trataremos, além disso, dalguns aspectos da análise de dados, sobretudo dalguns métodos de taxonomia numérica, os quais são indispensáveis para efectuar o agrupamento (classificação) dos referidos artefactos a partir dos resultados obtidos nas análises radioquímicas.

Tanto no primeiro caso como no segundo, teremos como objectivo informar os arqueólogos portugueses sobre os aspectos

fundamentais dos métodos que estamos a utilizar, de forma a que a colaboração que se iniciou possa vir a concretizar-se com a maior eficácia.

Há que notar, todavia, que o método de análise por activação é igualmente aplicável ao estudo de artefactos doutros materiais, como metais e rochas, sobretudo a obsidiana.

Note-se, além disso, que os métodos de taxonomia numérica são também usados para efectuar classificações e ordenações de artefactos com base em dados doutra natureza, como por exemplo resultados obtidos em observações tradicionais da forma e do estilo decorativo desses artefactos.

Espera-se, portanto, que esta informação possa contribuir paralelamente para que a colaboração de arqueólogos portugueses com investigadores portugueses doutras disciplinas venha a ampliar-se no futuro e a enveredar por novas vias de investigação em arqueologia, significativas no contexto do nosso território.

2. CARACTERIZAÇÃO DE CERÂMICAS MEDIANTE ANÁLISE INSTRUMENTAL POR ACTIVAÇÃO COM NEUTRÕES TÉRMICOS

2.1 —EFEITOS DA ACTIVAÇÃO COM NEUTRÕES TÉRMICOS

Quando se coloca uma amostra dum dado material durante algum tempo num reactor nuclear, que é a fonte mais usual de neutrões térmicos, muitos dos átomos dos elementos que compõem esse material experimentam reacções nucleares de absorção de neutrões e dão lugar a novos átomos.

Os átomos assim gerados são normalmente radioactivos, ou seja, os núcleos atómicos produzidos nessas reacções acabam por se transformar espontaneamente noutros núcleos, emitindo partículas elementares. Acontece, por outro lado, que os produtos resultantes das transformações espontâneas das espécies radioactivas são obtidos muitas vezes em estados nucleares excitados, isto é, com energias superiores à do seu estado fundamental (de energia mínima), os quais se desexcitam muito rapidamente

emitindo em geral radiação electromagnética, a que se dá o nome de raios gama, cuja energia é característica de cada espécie.

Assim, se o material que se coloca no reactor nuclear contiver sódio, por exemplo, alguns núcleos de átomos de sódio presentes na amostra absorvem um neutrão e dão lugar a ^{24}Na (sódio 24), que é uma espécie radioactiva. Esta espécie caracteriza-se pelo facto de na sua transformação espontânea dar origem a ^{24}Mg (magnésio 24) num estado nuclear excitado, que se desexcita imediatamente por emissão de dois raios gama, um de energia 1368 kev e outro de energia 2750 kev.

“Na desintegração \rightarrow $^{24}\text{Mg}^*$ desexcitação \rightarrow $^{24}\text{Mg} + \text{Raj(JS gama}$

sódio	-f partículas	magnésio
radioactivo	elementares	estável

As desintegrações radioactivas são processos de natureza aleatória. Se se considerar uma amostra onde, num determinado instante, haja uma certa quantidade de átomos duma dada espécie radioactiva, verifica-se que alguns deles se transformam ao fim dum curto intervalo de tempo, enquanto outros subsistem durante intervalos longos. É possível, todavia, definir uma vida média para esses átomos, a qual representa o tempo que eles durarão, em média, na amostra considerada.

A vida média dos átomos radioactivos também é característica de cada espécie. Para algumas espécies ela pode atingir muitos anos, para outras ela é apenas uma fracção ínfima de segundo. Por vezes, é conveniente exprimir a vida média duma dada espécie radioactiva em termos do período, definido como o tempo requerido para um número grande de átomos dessa espécie se reduzir a metade. O ^{24}Na , por exemplo, caracteriza-se por ter um período de 15 horas, a que corresponde uma vida média de cerca de 10 horas.

2.2 — FUNDAMENTO DOS MÉTODOS DE ANÁLISE, QUALITATIVA E QUANTITATIVA, POR ACTIVAÇÃO COM NEUTRÕES TÉRMICOS

A análise qualitativa dum dado material por activação com neutrões térmicos consiste, pois, em irradiar uma amostra desse material numa certa posição dum reactor nuclear, ou de qualquer outra fonte de neutrões, e em medir as energias dos raios gama emitidos pela amostra irradiada, com auxílio dum espectrómetro de raios gama.

Se, por exemplo, se verificar que alguns desses raios têm energia de 1368 kev, poderá então concluir-se com toda a segurança que o material contém sódio.

A identificação dos elementos constituintes dos materiais, embora necessária, não é contudo suficiente para caracterizar esses materiais. Torna-se necessário também determinar a sua concentração, isto é, a sua quantidade por unidade de massa de material.

A análise quantitativa realiza-se, neste caso, medindo as taxas de contagem das espécies radioactivas produzidas a partir dos referidos elementos, as quais são proporcionais ao número de transformações espontâneas verificadas para essas espécies por unidade de tempo, ou seja, às suas actividades.

A actividade dum dado espécie radioactiva, produzida irradiando a amostra na fonte de neutrões, é directamente proporcional à massa do elemento que lhe dá origem. Assim, quanto maior for a quantidade de elemento na amostra, tanto maior será, para o mesmo tempo e as mesmas condições de irradiação, a actividade da espécie radioactiva a que esse elemento dá lugar e, por isso, tanto maior será a sua taxa de contagem para as mesmas condições de medição.

Na prática o que se costuma fazer é irradiar juntamente com a amostra um padrão do elemento que se deseja analisar (onde se conhece a sua concentração) e comparar a taxa de contagem da espécie radioactiva produzida na amostra com a taxa de contagem da mesma espécie produzida no padrão, medidas

nas mesmas condições experimentais. A massa M desse elemento na amostra obtém-se, então, a partir da expressão:

$$M = M' \cdot \frac{T \text{ (amostra)}}{T \text{ (padrão)}} \quad (1)$$

onde M' representa o valor conhecido da massa do mesmo elemento no padrão e as letras T designam taxas de contagem.

Dado que a taxa de contagem da espécie radioactiva na amostra não pode ser medida no mesmo momento em que se mede a taxa de contagem da mesma espécie no padrão, mas sim antes ou depois dum certo intervalo de tempo, há que fazer a sua correcção recorrendo à lei do declínio radioactivo, de modo a referi-las a um instante comum, em particular quando se trata de espécies radioactivas de pequena vida média. A expressão (1) só é válida, portanto, uma vez efectuada essa correcção.

2.3 — MÉTODO INSTRUMENTAL POR ACTIVAÇÃO COM NEUTRÕES TÉRMICOS

Como vimos, qualquer método de análise por activação com neutrões térmicos compreende duas fases principais: (1) a irradiação da amostra e do padrão numa fonte de neutrões; (2) a medição das taxas de contagem da espécie radioactiva produzida a partir do elemento a analisar, tanto na amostra como no padrão. Por vezes, torna-se indispensável introduzir uma fase intermédia, envolvendo a destruição da amostra e mais tratamentos químicos, sempre que na medição da taxa de contagem da espécie radioactiva em causa haja interferência doutras espécies radioactivas, produzidas simultaneamente a partir doutros elementos também existentes na amostra. Isso sucede, por via de regra, quando não se dispõe de espectrómetros de raios gama de elevada resolução, ou seja, capazes de permitir distinguir raios gama de energias muito próximas.

Assim, é costume classificar os métodos de análise por activação com neutrões térmicos em dois grandes grupos: (a) métodos destrutivos, caracterizados pela necessidade de se solubilizar a amostra e de se efectuar a separação química das espécies radioactivas obtidas na irradiação a partir dos elementos, a analisar; (b) métodos instrumentais, ou seja, aqueles em que essa solubilização e separação são dispensadas através do emprego dum espectrómetro de raios gama.

O método de análise por activação com neutrões térmicos que está a ser empregado no LFEN para caracterizar cerâmicas arqueológicas é, como se disse, um método instrumental.

2.4 — PREPARAÇÃO DE AMOSTRAS E PADRÕES

Os fragmentos de cerâmicas recebidos no laboratório são, primeiramente, limpos em determinadas regiões com um escopro de ponta de carboneto de tungsténio, a fim de se remover das superfícies dessas regiões quaisquer materiais estranhos que possam estar-lhes agregados. Em seguida, a partir das superfícies limpas, abrem-se furos usando uma broca também com ponta de carboneto de tungsténio. Recolhe-se, assim, cerca de 1 g de cerâmica em pó.

Esta porção é, depois, introduzida num recipiente dum misturador de sólidos onde é agitada e, uma vez terminada essa operação, é levada para uma estufa e seca. Seguidamente, pesa-se uma certa quantidade do material seco (da ordem de grandeza de 100 mg) e introduz-se essa amostra numa pequena ampola de quartzo, a qual é então fechada ao maçarico e, por fim, limpa cuidadosamente.

O quartzo utilizado tem uma pureza excepcionalmente elevada, a fim de permitir que as medições das taxas de contagem, após a irradiação no reactor nuclear, possam ser efectuadas sem retirar as amostras das ampolas, o que simplifica muitíssimo o trabalho.

Por vezes, em particular nos casos de amostras que só têm de ser irradiadas durante um pequeno intervalo de tempo, o quartzo é substituído por polietileno.

Os padrões que estão a ser utilizados são rochas preparadas na instituição americana «U. S. Geological Survey», designadamente as seguintes: G-2, GSP-1, AGV-1, BCR-1, PCC-1 e DTS-1. Estes padrões já foram analisados em numerosos laboratórios e conhecem-se, por isso, as concentrações dos seus elementos constituintes, cujos valores mais prováveis foram tabelados por Flanagan¹² e Harbottle e colaboradores¹³. Os valores que estamos a usar são os publicados por Harbottle.

Os padrões são igualmente secos e, uma vez secos, manuseados da mesma maneira que as amostras.

2.5 — IRRADIAÇÃO DE AMOSTRAS E PADRÕES NO REACTOR NUCLEAR

As irradiações de amostras e padrões fazem-se no reactor português de investigação (RPI), do Laboratório de Sacavém, em posições não muito próximas do núcleo, a fim de se reduzir a acção interferente de neutrões rápidos. Nessas posições o fluxo de neutrões térmicos é de cerca de 2×10^{12} neutrões por cm^2 e por segundo.

Para efectuar uma análise completa, amostras e padrões são submetidos a duas irradiações: a primeira, de 15 minutos, com o objectivo de se determinar a concentração de elementos que dão lugar a espécies radioactivas de pequena vida média, em particular de sódio, potássio, manganês, lantânio e disprósio; a segunda, de 5 dias, intermitente, dado que o reactor só funciona 14 horas por dia, tendo em vista a determinação da concentração de elementos que conduzem à produção de espécies radioactivas de vida média superior a 1 dia, em particular de mais 14 elementos, nomeadamente os seguintes: ferro, rubídio, cézio, bário, cobalto, escândio, európio, lutécio, crómio, cério, tório, háfnio, antimónio e tántalo. No Quadro 1 indicam-se algumas características das espécies radioactivas obtidas a partir dos mencionados elementos.

QUADRO I

CARACTERÍSTICAS DAS ESPÉCIES RADIOACTIVAS QUE SERVEM DE BASE À ANÁLISE POR ACTIVAÇÃO DE ELEMENTOS CONSTITUINTES DAS CERÂMICAS

Elemento	Espécie radioactiva	Periodo da espécie radioactiva	Energia dos raios gama seleccionados (kev)
Disprósio	¹⁶⁵ Dy	2,36 h	95
Manganês	⁵⁶ Mn	2,58 h	847
Sódio	²⁴ Na	15,00 h	1368
Potássio	⁴² K	12,42 h	1525
Lantânio	¹⁴⁰ La	40,22 h	1596
Cério	¹⁴¹ Ce	32,53 d	145
Lutécio	¹⁷⁷ Lu	6,72 d	208
Tório	²³³ Pa	27,00 d	312
Crómio	⁵¹ Cr	27,80 d	320
Háfnio	¹⁸¹ Hf	42,50 d	482
Bário	¹³¹ Ba	12,00 d	496
Césio	¹³⁴ Cs	2,05 a	796
Escândio	⁴⁸ Sc	83,90 d	889
Rubídio	⁸⁶ Rb	18,66 d	1077
Tântalo	¹⁸² Ta	115,00 d	1221
Ferro	⁵⁹ Fe	45,00 d	1291
Cobalto	⁶⁰ Co	5,26 a	1332
Európio	¹⁵² Eu	12,00 a	1408
Antimónio	¹²⁴ Sb	60,30 d	1691

Note-se que dos elementos que se analisam só três, o sódio, o potássio e o ferro, são constituintes principais das cerâmicas, isto é, estão presentes em concentrações da ordem de grandeza de 1% a 10%, em peso. Os restantes são constituintes menores, quer dizer, as suas concentrações são da ordem de grandeza de 1 a 100 partes por milhão. Apesar disso, a precisão com que se determinam as concentrações de muitos dos elementos menores é tão boa como a obtida na determinação das concentrações de sódio, potássio e ferro. Esta circunstância deve-se ao facto de a sensibilidade dos métodos de análise por activação com neutrões tér-

micos na determinação da concentração dum dado elemento depender sobretudo de dois factores: do fluxo de neutrões térmicos a que é exposta a amostra a analisar; e da probabilidade de ocorrência da reacção nuclear que dá origem à espécie radioactiva cuja taxa de contagem tem de medir-se. Por outras palavras, há elementos que, embora presentes em grande percentagem, podem dar lugar a quantidades muito pequenas de espécies radioactivas. Há outros que, pelo contrário, mesmo existindo em concentrações muitíssimo pequenas, podem dar origem, sob o mesmo fluxo de neutrões, a quantidades muito maiores de espécies radioactivas. Isso depende, fundamentalmente, da sua probabilidade de reagirem com os neutrões.

O período da espécie radioactiva pouca influência exerce na sensibilidade do método. Impõe, porém, certas limitações de ordem prática nos casos em que é grande ou muito pequeno: no primeiro caso, devido à necessidade de se irradiar a amostra durante muito tempo; no segundo, por se tornar muito difícil medir a taxa de contagem da espécie radioactiva. Determina, por outro lado, os instantes mais adequados para efectuar as medições de taxas de contagem.

2.6 — ESPECTROMETRIA DE RAIOS GAMA

As medições de taxas de contagem estão a ser feitas recorrendo a um espectrómetro de raios gama com um detector de Ge(Li) de elevada resolução. O detector está ligado, através de unidades de amplificação, a um analisador de impulsos de 4000 canais, que está por sua vez associado a uma perfuradora de fita, destinada a registar os espectros em fita perfurada.

Fazem-se quatro séries de medições. A primeira inicia-se 2 horas depois de terminada a irradiação curta, a fim de se medirem as taxas de contagem das espécies radioactivas produzidas nas reacções nucleares com o disprósio e o manganês. O tempo de cada uma destas medições é de 10 minutos. A segunda é começada cerca de 20 horas depois de concluída a mesma irradiação, com o objectivo de se medirem as taxas de contagem das espécies radioactivas resultantes do sódio, potássio e lantânio. O tempo de cada medição é agora de 30 minutos.

A terceira série tem início 10 dias após o termo da irradiação longa e destina-se a medir a taxa de contagem da espécie radioactiva devida ao lutécio. Cada amostra é medida durante 30 minutos. A quarta começa-se 3 a 4 semanas depois de acabada esta mesma irradiação e tem por finalidade medir as taxas de contagem das espécies radioactivas produzidas pelos restantes 13 elementos. O tempo de cada medição é então de 90 minutos. A título de ilustração mostra-se na Fig. 1, sob forma gráfica, o espectro obtido neste último caso com uma amostra dum fragmento de cerâmica escura, da época pré-romana, encontrado em Santa Olaia.

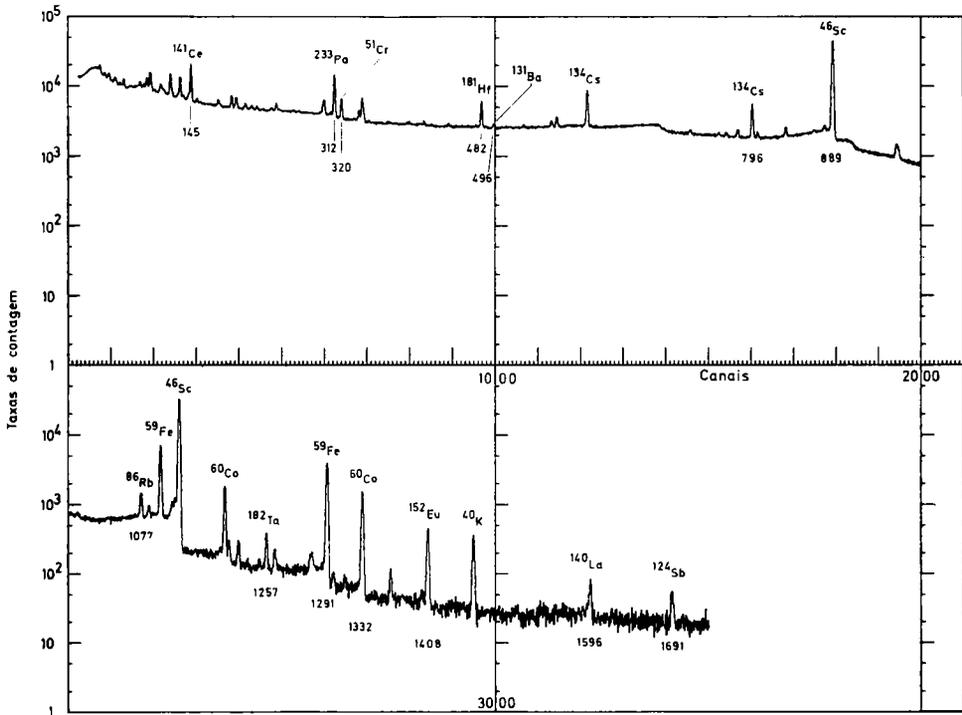


FIG. 1 — Espectro de raios gama duma cerâmica pré-romana, encontrada em Santa Olaia, tirado 26 dias depois de irradiada no reactor nuclear.

Note-se que cada espectro de raios gama debitado pelo analisador de 4000 canais é caracterizado por cerca de 24000 dígitos, os quais dificultam pela sua quantidade os cálculos numéricos que têm de efectuar-se para se chegar aos valores finais das taxas de contagem. Torna-se muito vantajoso, por conseguinte, utilizar computadores para fazer esses cálculos.

No LFEN recorre-se a um pequeno computador PDP 15, o único que existe no Laboratório, o qual não está ligado ao espectrómetro de raios gama. Assim, os espectros são introduzidos no computador através duma unidade de leitura de fitas perfuradas. Em seguida fazem-se correr programas apropriados, cuja função é não só efectuar a análise desses espectros e calcular os valores finais das taxas de contagem das diferentes espécies radioactivas, tendo em conta os tempos de medição e os tempos de declínio, mas também comparar os valores calculados para as amostras com os correspondentes valores calculados para os padrões e, através dessa comparação e do conhecimento dos valores das concentrações dos elementos nos padrões, determinar as concentrações dos vários elementos nas amostras.

O LFEN possui vários programas de análise de espectros de raios gama. Para espectros obtidos em análise por activação está-se a recorrer actualmente aos programas GELIAN¹⁴ e OLIVE¹⁵.

Espera-se, contudo, que estas condições experimentais, por enquanto precárias no domínio do registo e leitura de espectros, venham a ser melhoradas num futuro próximo, substituindo o registo em fita perfurada (muito lento) por um registo em fita magnética. Espera-se, por outro lado, acelerar a análise de espectros efectuando-a num computador mais rápido que o do Laboratório, como por exemplo no computador IBM 360/44 do Centro de Cálculo das Universidades de Lisboa.

2.7—ATRIBUTOS DAS CERÂMICAS ANALISADAS. ESTADOS DESSES ATRIBUTOS

Chegados ao termo da informação sobre os aspectos fundamentais do método instrumental por activação com neutrões térmicos e da sua aplicação à análise de cerâmicas, será lícito

que se pergunte para que vão servir afinal as análises que estão a realizar-se. A resposta é simples. Essas análises têm por objecto a caracterização das cerâmicas, ou seja, visam exactamente o mesmo objectivo que o visado pelo arqueólogo nas suas observações tradicionais.

A diferença entre o nosso processo e o tradicional está em que, em vez de se olhar para a sua cor, de se examinar o tipo de pasta, a forma do perfil, o pormenor da decoração, etc., se observam elementos químicos e se medem as suas concentrações. A aplicação do método instrumental por activação com neutrões térmicos na análise de cerâmicas vai permitir-nos, portanto, caracterizá-las por meio de 19 atributos (os elementos analisados) e dos seus correspondentes estados (as concentrações desses elementos).

A questão que se nos põe seguidamente é a mesma que se apresenta ao arqueólogo após as suas observações tradicionais, ou seja, a de agrupar as cerâmicas analisadas em grupos (classes) com base nos estados desses atributos. Dela nos ocuparemos na alínea 3 do presente artigo.

Não deve esquecer-se, porém, que uma grande parte do que vai dizer-se a esse respeito se aplica igualmente ao agrupamento de artefactos doutros materiais, quer a sua caracterização seja feita recorrendo aos mesmos atributos, quer a quaisquer outros, incluindo os relacionados com a forma e o estilo decorativo, os quais têm normalmente carácter qualitativo.

3. ANÁLISE DE DADOS

O agrupamento (classificação) de quaisquer entidades fundamenta-se na apreciação das suas semelhanças (ou dissemelhanças) relativas. O primeiro problema que surge a quem tenha de efectuar agrupamentos é, por isso, o de determinar essas semelhanças (ou dissemelhanças), com base nos atributos que forem considerados.

Começaremos, pois, por discutir muito sumariamente alguns métodos numéricos de fazer essa determinação.

3.1 — MATRIZ DE DADOS

Para se calcular a semelhança (ou dissemelhança) entre pares de entidades, geralmente designadas em taxonomia numérica por unidades taxonómicas operacionais (OTU's), é costume dispor os dados sob a forma duma matriz de t colunas e n linhas, em que as t colunas representam os t OTU's a agrupar e as n linhas os n atributos usados na caracterização desses OTU's. No Quadro 2 mostra-se simbolicamente como se faz esta arrumação.

QUADRO 2
MATRIZ DE DADOS

Atributos	OTU's			
	1	2	j	t
1	X_{11}	X_{12}	X_{1j}	X_{1t}
2	X_{21}	X_{22}	X_{2j}	X_{2t}
⋮	⋮	⋮	⋮	⋮
i	X_{i1}	X_{i2}	X_{ij}	X_{it}
⋮	⋮	⋮	⋮	⋮
n	X_{n1}	X_{n2}	X_{nj}	X_{nt}

Nesta matriz de dados X_{ij} representa, por conseguinte, o estado do atributo i no OTU j . Assim, no nosso caso particular $n = 19$ e, se por exemplo j for a amostra duma das t cerâmicas analisadas e i o atributo sódio, será o valor da concentração de sódio nessa amostra.

8.2 — OPERAÇÕES DE TRANSFORMAÇÃO DOS DADOS

Um dos princípios fundamentais da taxonomia numérica¹⁶ consiste em admitir que, a priori, todos os atributos deverão ter peso igual na formação de grupos, ou seja, em considerar que não

se deve dar a nenhum atributo maior importância do que aos outros.

Ora verifica-se muitas vezes que os estados dum certo atributo (distribuídos ao longo duma certa linha da matriz de dados), usado para caracterizar os vários OTU's, variam por vezes entre limites afastados. No caso dos elementos analisados nas cerâmicas, por activação com neutrões térmicos, esse afastamento é, em geral, muito pouco acentuado. Contudo, tanto para o sódio como para o manganês, observam-se ainda, de cerâmica para cerâmica, variações apreciáveis nas suas concentrações que chegam a atingir uma ordem de grandeza, isto é, verifica-se que as suas concentrações nalgumas cerâmicas são dez vezes superiores às encontradas noutras. Considera-se, por isso, vantajoso efectuar transformações nos dados, no sentido de que cada atributo contribua para os coeficientes de semelhança (ou dissemelhança) em proporção inversa com a extensão da variação dos seus estados nos diversos OTU's, ou seja, no sentido de igualar tanto quanto possível essa variação.

Um dos processos mais correntes de fazer tais transformações é o da estandardização de dados¹⁷, o qual consiste em calcular a média X e o desvio padrão s de cada atributo (dos valores de cada linha da matriz de dados) e em substituir cada valor pelo seu desvio $(X-X)$ relativamente à média, expresso em unidades de desvio padrão, isto é, em substituir cada valor de X por $(X-X)/s$.

Outro processo¹⁸, mais simples que o anterior, consiste em calcular o intervalo de variação $(X_{max}-X_{min})$ de cada atributo e em substituir cada valor pelo desvio $(X-X_{min})/(X_{max}-X_{min})$ relativamente ao valor mínimo, expresso em unidades de intervalo de variação, ou seja, em substituir cada valor de X por $(X-X_{min})/(X_{max}-X_{min})$. Alguns investigadores chamam-lhe processo do intervalo de variação, outros processo de normalização.

Um terceiro processo¹⁹ consiste em aplicar aos dados o operador logaritmo, isto é, em substituir os dados originais pelos respectivos logaritmos. Há ainda quem associe este processo ao da estandardização, atrás referido, aplicando o operador logaritmo antes de proceder à estandardização¹⁹. Deve notar-se, porém, que a transformação logarítmica é contestada por alguns investigadores.

Contudo, essa contestação é mais de ordem fundamental e, na prática, os grupos obtidos com base em dados transformados parecem não diferir apreciavelmente com o processo usado na transformação.

Assim, na determinação da semelhança (ou dissemelhança) entre pares de OTU's parte-se, por via de regra, duma matriz de dados transformados, a qual tem obviamente o mesmo número de colunas e o mesmo número de linhas que a matriz de dados originais.

B.3 — COEFICIENTES DE SEMELHANÇA

A semelhança (ou a dissemelhança) entre dois OTU 's é normalmente determinada em termos dum coeficiente que a quantifica, a que se dá o nome de coeficiente de semelhança (ou dissemelhança).

Conhecem-se vários tipos destes coeficientes, os quais são geralmente divididos em quatro grupos: (1) distâncias; (2) coeficientes de correlação; (3) coeficientes de associação e (4) coeficientes probabilísticos. Aqui, apenas nos ocuparemos dalguns incluídos nos dois primeiros grupos, por serem os mais utilizados quando se dispõe de dados quantitativos e os que iremos aplicar no nosso projecto de investigação de cerâmicas. Deve notar-se, todavia, que os coeficientes de associação, empregados sobretudo quando alguns ou todos os atributos são de natureza qualitativa (forma e pormenores de decoração, por exemplo), têm sido também muito usados em estudos de arqueologia.

3.8.1 — *Distâncias*

Existem vários coeficientes de dissemelhança deste tipo. A distância euclidiana é um dos mais empregados e o mais facilmente perceptível.

Consideremos, por exemplo, dois OTU's j e k e suponhamos que se podem caracterizar medindo em cada um os estados de dois atributos, que designaremos por 1 e 2. Se representarmos esses OTU's por pontos num espaço a 2 dimensões, usando como coordenadas os estados desses atributos, (X_{i-}, X_{2j}) para o OTU j

e (X_{1j}, X_{2j}) para o OTU k (ver Fig. 2), a distância euclidiana entre os OTU's j e k , obtém-se facilmente aplicando o bem conhecido teorema de Pitágoras:

$$\Delta_{jk} = [(X_{1j} - X_{1k})^2 + (X_{2j} - X_{2k})^2]^{1/2} \quad (2)$$

Se os tivéssemos caracterizado medindo em cada um os estados de três atributos (1, 2 e 3) e representado por pontos num espaço a 3 dimensões, utilizando como coordenadas os estados desses atributos, (X_{1j}, X_{2j}, X_{3j}) para o OTU j e (X_{1k}, X_{2k}, X_{3k}) para o OTU k (ver Fig. 3), a distância euclidiana entre os referidos OTU's seria:

$$\Delta_{jk} = [(X_{1j} - X_{1k})^2 + (X_{2j} - X_{2k})^2 + (X_{3j} - X_{3k})^2]^{1/2} \quad (3)$$

Podemos generalizar este conceito de distância entre dois pontos num espaço a n dimensões (tantas quantos os atributos usados para caracterizar os OTU's) e, então, a distância euclidiana será dada por:

$$\Delta_{jk} = [(X_{1j} - X_{1k})^2 + (X_{2j} - X_{2k})^2 + \dots + (X_{nj} - X_{nk})^2]^{1/2}$$

ou

$$\Delta_{jk} = \left[\sum_{i=1}^n (X_{ij} - X_{ik})^2 \right]^{1/2}$$

onde X_{ij} representa, como se viu atrás, o estado do atributo i no OTU j e X_{ik} o estado do mesmo atributo no OTU k .

Como n aumenta com o número de atributos usados na caracterização dos OTU's, a distância que geralmente se calcula é uma distância média, d , denominada distância euclidiana média e definida por:

$$d_{jk} = \left[\frac{1}{n} \sum_{i=1}^n (X_{ij} - X_{ik})^2 \right]^{1/2}$$

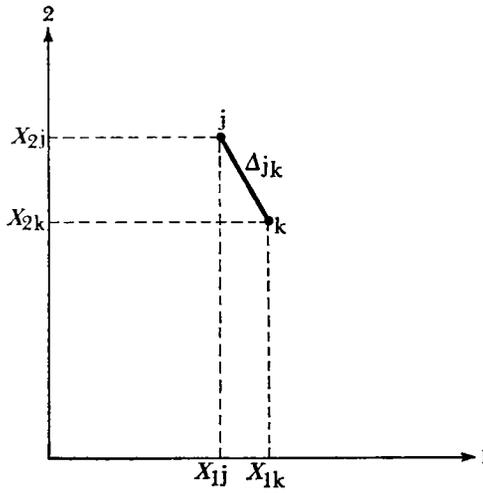


FIG. 2 — Ilustração do conceito de distância euclidiana entre dois OTU's j e k , num espaço a 2 dimensões.

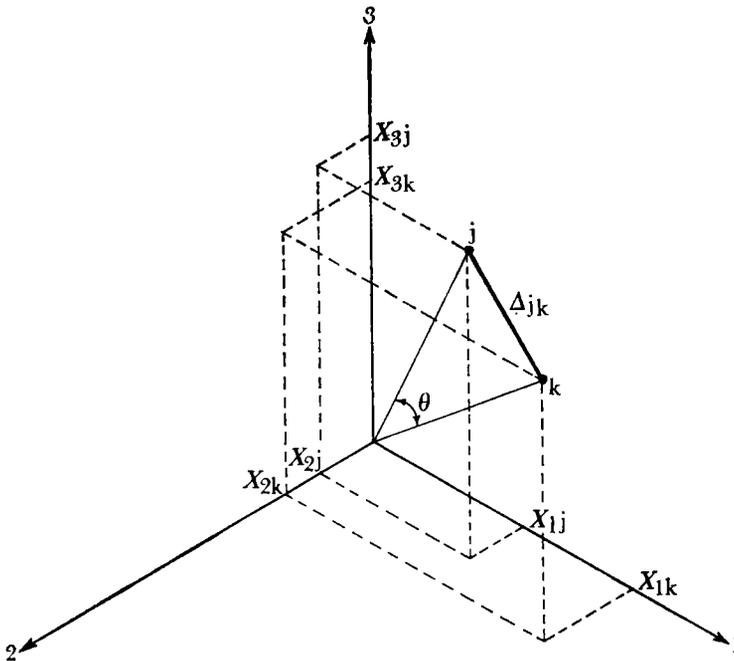


FIG. 3 — Ilustração do conceito de distância euclidiana entre dois OTU's j e k , num espaço a 3 dimensões.

Repare-se que, se os OTU's j e k forem idênticos, os estados de cada atributo nos dois OTU's serão iguais ($X_{ij} = X_{ik}$) e a distância euclidiana média entre eles será nula. Se eles forem diferentes, d_{jk} será maior que zero e tanto maior quanto maior for a sua diferença.

Outro conceito de distância que tem sido muito aplicado é o de distância Manhattan média, definida por:

$$(MCD)_{jk} = \frac{1}{n} \sum_{i=1}^n |X_{ij} - X_{ik}| \quad (6)$$

A esta distância dá-se também o nome de «diferença média entre atributos» e daí o representar-se usualmente pelas letras MCD, que são as iniciais das palavras inglesas «mean character difference». Tal como anteriormente, se j e k forem idênticos $(MCD)_{jk} = 0$; se forem diferentes $(MCD)_{jk} > 0$.

3.3.2 — Coeficientes de correlação

Um dos coeficientes de correlação mais empregados é o de Pearson²⁰, o qual, referido a dois OTU's j e k , se define do seguinte modo:

$$r_{jk} = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{\left[\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 \sum_{i=1}^n (X_{ik} - \bar{X}_k)^2 \right]^{1/2}} \quad (7)$$

onde X_{ij} e X_{ik} têm o mesmo significado que o exposto na alínea anterior e \bar{X}_j e \bar{X}_k são as médias dos valores dos estados dos atributos do OTU j e do OTU k , respectivamente.

Para dados estandardizados este coeficiente é aproximadamente igual ao coseno do ângulo formado pelos segmentos de recta que unem esses pontos com a origem dos eixos coordenados²¹. Na representação no espaço a 3 dimensões da Fig. 3 esse ângulo foi designado por θ .

Note-se que, se os OTU's j e k forem idênticos, $0 = 0$, $\cos 0 = 1$ e $T_{jkc} = 1$. Se não forem, $0 \neq 0$ e $r^{\wedge} < 1$. Pode acontecer, no entanto, que j não seja idêntico a k mas que, apesar disso, seja $T_{jkc} = 1$. É o caso de dois OTU's representados por dois pontos, afastados um do outro, mas situados no mesmo prolongamento em relação à origem dos eixos coordenados. Isso significará que, embora não idênticos, deve existir entre eles uma correlação íntima.

8.8.3 — Matriz de semelhanças

Os valores calculados dos coeficientes de semelhança (ou dissemelhança) respeitantes aos diversos pares que os OTU's podem constituir, formam uma matriz com t colunas e t linhas (tantas quantos os OTU's considerados), à qual se dá o nome de matriz de semelhanças (ou dissemelhanças). A título de exemplo apresenta-se no Quadro 3 uma matriz de dissemelhanças, relativa ao caso em que se usa como coeficiente a distância euclidiana média.

QUADRO 3

MATRIZ DE DISTÂNCIAS

OTU's	OTU's			
	1	2	3	t
1	0	d_{12}	d_{13}	d_{1t}
2	d_{21}	0	d_{23}	d_{2t}
3	d_{31}	d_{32}	0	d_{3t}
⋮	⋮	⋮	⋮	⋮
t	d_{t1}	d_{t2}	d_{t3}	0

A ordem pela qual são referidos os OTU's por colunas é a mesma pela qual se referem por linhas e, por isso, cada valor da diagonal principal desta matriz representa a dissemelhança de

cada OTU consigo próprio, isto é, representa a identidade. Como a matriz indicada diz respeito a distâncias a sua diagonal principal é constituída por zeros. Seria constituída por uns se dissesse respeito a coeficientes de correlação.

Repare-se que, como a dissemelhança dum dado OTU j com um OTU k é a mesma que a de k com j ($d^j = d^k$), o triângulo superior direito daquela matriz (situado acima dos zeros) é igual à imagem do triângulo inferior esquerdo (situado abaixo dos zeros) — a matriz é simétrica — e, portanto, ela pode ser representada apenas por uma das partes. Usualmente só se apresenta a correspondente ao triângulo inferior esquerdo.

É importante reconhecer que os valores numéricos dos coeficientes de semelhança (ou dissemelhança) calculados a partir das expressões atrás apontadas não têm carácter absoluto e que apenas devem ser comparados no âmbito duma certa classificação.

8.4 _ ANÁLISE DE GRUPOS

O conceito de grupo é difícil de definir, devido em grande parte à existência dum considerável número de critérios para medir as suas propriedades. Por isso, muitos investigadores, como Sneath e Sokal¹⁶, preferem referir-se-lhe duma maneira vaga, considerando-o como um conjunto de OTU's, num certo hiperespaço, que não apresenta características duma distribuição aleatória nem regular e que satisfaz a um ou mais critérios impostos por dadas definições particulares.

Ainda segundo esses investigadores, uma das maneiras de caracterizar um grupo é através da sua densidade (número de OTU's por unidade de hipervolume), a qual deve ser maior do que em outras regiões do espaço considerado.

Assim, diremos, usando também uma maneira vaga de expressão, que a análise de grupos, cujo objectivo é o agrupamento (a classificação) de OTU's, consiste em procurar as regiões do hiperespaço considerado (no caso presente o espaço de atributos) a que correspondem densidades apreciáveis de OTU's.

Essa análise pode ser efectuada recorrendo a vários métodos, dos quais se distinguem diferentes tipos: de aglomeração e partição;

hierárquicos e não-hierárquicos; de sobreposição e não-sobreposição; sequenciais e simultâneos; directos e iterativos; etc.. O leitor poderá encontrar no livro daqueles investigadores uma explicação clara das suas características.

Nas páginas seguintes vamos apenas ocupar-nos, muito brevemente, dalguns dos métodos mais utilizados e que se pensa vir a aplicar no agrupamento das cerâmicas examinadas no nosso projecto de investigação.

3.4.1 — *Métodos SAHN*

A designação SAHN representa o conjunto de iniciais das palavras sequencial, aglomerativo, hierárquico e não-sobreposição, que qualificam os referidos métodos. Sequencial quer dizer que se aplica uma sequência de operações repetitivas aos diversos OTU's, quer considerados individualmente, quer como se constituíssem um conjunto unitário. Aglomerativo indica que, nessa aplicação, se principia com t OTU's e se vai agrupando sucessivamente esses OTU's. Hierárquico significa que se efectua uma sequência de agrupamentos ao longo da qual o número de grupos vai consequentemente diminuindo, mas tal que, para um dado nível de agrupamento, não há limitação no grau de sobreposição dos grupos obtidos, isto é, que um OTU pode ser simultaneamente membro de dois ou mais grupos. Não-sobreposição quer dizer que, para qualquer nível de agrupamento, um OTU só pode ser membro dum único grupo.

Conhecem-se vários métodos SAHN de efectuar a análise de grupos, os quais diferem sobretudo no critério utilizado para fazer os sucessivos agrupamentos. A fim de não se tornar demasiado longa a exposição, referir-nos-emos apenas a dois dos métodos mais frequentemente usados em estudos de arqueologia: o método UPGMA e o da minimização do aumento da dimensão de grupos.

3.4.1.1 — *Método UPGMA*

A denominação UPGMA deriva das iniciais das palavras inglesas empregadas para o designar, nomeadamente as seguintes: «unweighted pair-group method using arithmetic averages».

Neste método, quando aplicado por exemplo à matriz de distâncias, anteriormente referida, recorre-se ao conceito de distância entre dois grupos J e K . d_{JK} , definida pela expressão:

$$d_{JK} = \frac{1}{t_J t_K} \sum d_{jk} \quad (8)$$

onde t_J e t_K são o número de OTU's no grupo J e o número de OTU's no grupo K , respectivamente, e o somatório das distâncias d_{jk} , entre OTU's, se estende a todos os pares de OTU's que é possível formar com um OTU de J e um OTU de K .

Tal como noutros métodos do tipo SAHN, começa-se por admitir que no conjunto de OTU's considerados existem inicialmente tantos grupos quantos os OTU's, cujas distâncias (dissimilaridades) entre si são as que constituem a matriz de distâncias atrás apresentada.

Seleccionam-se, então, numa primeira fase, os pares mutuamente mais semelhantes e considera-se que cada um desses pares passa a formar um grupo. Depois, calculam-se por intermédio da expressão (8) as distâncias entre grupos tomando em conta as fusões efectuadas, seleccionam-se os pares mutuamente mais semelhantes e funde-se cada um desses pares. Esta segunda fase é seguida por outras idênticas, ou seja, em cada fase sucessiva de agrupamentos calculam-se as distâncias entre os grupos existentes ao terminar a fase anterior, seleccionam-se os pares mutuamente mais semelhantes e funde-se cada um deles. A última fase consiste em reunir num único grupo o par a que se chega no penúltimo.

3.4.1.2 — Método da minimização do aumento da dimensão de grupos

Neste método os sucessivos agrupamentos são feitos com base nas definições de dimensão de grupo e de distância entre dois grupos a seguir indicadas.

A dimensão, S_j , dum dado grupo j , é definida pela expressão :

$$S_j = \frac{1}{t_j} \sum d_{jk} \quad (9)$$

onde t_j representa o número de OTU's no grupo j e o somatório das distâncias d se estende a todos os pares de OTU's que é possível formar com os constituintes desse grupo.

A distância entre dois grupos J e K , d_{JK} é agora definida como o aumento de dimensão verificado ao reunir esses grupos num só, que designaremos por R .

$$d_{JK} = S_R - (S_J + S_K) \quad (10)$$

O método pratica-se de maneira idêntica à descrita anteriormente. Parte-se da hipótese de que no conjunto de OTU's considerados existem inicialmente tantos grupos quantos os OTU's, cujas distâncias entre si são as incluídas na matriz de distâncias. Funde-se, então, numa primeira fase, cada par de grupos mutuamente mais semelhantes. Em seguida, calculam-se as distâncias entre grupos entrando em consideração com os agrupamentos feitos, mas recorrendo agora às expressões (9) e (10), e reúne-se cada par de grupos mutuamente mais semelhantes. Repetem-se depois as operações de agrupamento, sucessivamente, tomando sempre por base em cada fase os agrupamentos efectuados na fase anterior, até se chegar a um único grupo.

3.4.1.8 — Representação de resultados

Os resultados da análise de grupos são usualmente representados sob a forma de gráficos, a que se dá o nome de dendrogramas. Alguns investigadores distinguem dois tipos: fenogramas, que representam relações fenéticas (semelhanças); e cladogramas, que traduzem sequências evolucionárias. Na representação dos nossos resultados empregaremos, portanto, fenogramas.

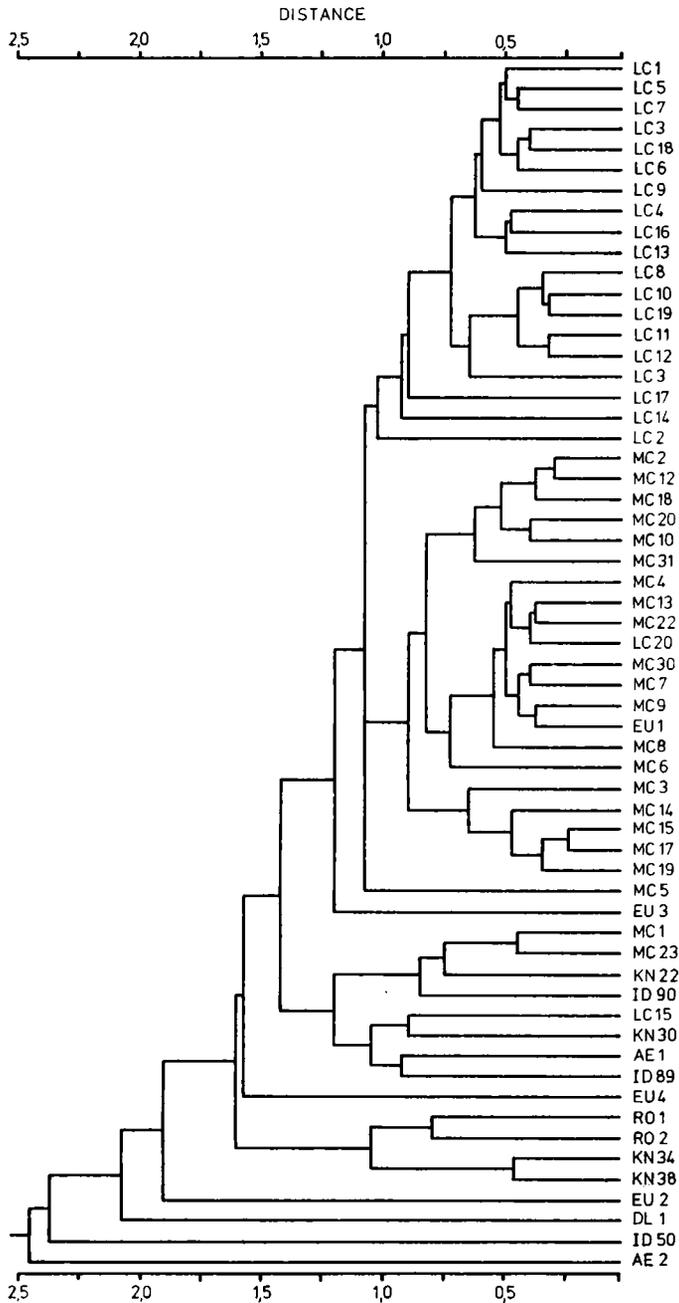


Fig. 4 — Fenograma representativo do agrupamento de 57 cerâmicas gregas, obtido caracterizando-as pelas concentrações de 17 elementos, calculando distâncias euclidianas médias e aplicando o método UPGMA. Os símbolos referem-se aos lugares onde as cerâmicas foram encontradas: LC, Lacônia; MC, Micenas; KN, Cnossos; AE, Egina; DL; Delfos; EU, Eutresis; ID, Idália; RO, Rodas.

Os fenogramas são diagramas ramificados, onde os níveis a que se unem os diferentes ramos estão relacionados com os valores dos coeficientes de semelhança (ou dissemelhança) em que se baseia o método utilizado de análise de grupos. A título de exemplo, mostra-se na Fig. 4 o fenograma representativo dos resultados obtidos num estudo²² de classificação de cerâmicas gregas, caracterizadas pelas concentrações de 17 elementos analisados por activação com neutrões térmicos²³, efectuado usando como coeficiente de dissemelhança a distância euclidiana média, expressa pela equação (5), e como método de análise de grupos o método UPGMA. Há, todavia, quem apresente os fenogramas numa maneira um pouco diferente, com alguns ramos cortados, tal como se mostra na Fig. 5 que representa os resultados dum estudo de cerâmicas Maia e de argilas de Lubaantun²⁴, também efectuado caracterizando as cerâmicas pelas concentrações de 19 elementos analisados por activação com neutrões térmicos, determinando numericamente as suas dissemelhanças por intermédio da equação (5) e fazendo a análise de grupos pelo método UPGMA.

O grau de ajustamento dum fenograma à correspondente matriz de semelhanças (ou dissemelhanças) pode ser calculado sob a forma dum coeficiente, a que se chama coeficiente de correlação cofenética²⁵. Trata-se do coeficiente de correlação entre os valores de mínima dissemelhança (ou de máxima semelhança), associados a cada par de OTU's e expressos no fenograma, os chamados valores cofenéticos, e os respectivos valores da matriz de dissemelhanças (ou semelhanças). O valor do coeficiente de correlação cofenética varia normalmente entre 0,60 e 0,95, sendo tanto maior quanto melhor for esse ajustamento. Nos estudos de cerâmicas arqueológicas, que têm sido efectuados, tem-se verificado uma variação entre 0,60 e 0,80¹⁹.

3.4.2 — Métodos das *k*-médias

Estes métodos de análise de grupos são métodos de partição? não-hierárquicos, que consistem, resumidamente, em repartir o conjunto de *t* OTU's (considerado unitário) em *k* grupos, cada um dos quais é caracterizado pelas médias dos estados dos atributos dos seus constituintes, ou seja, pelo respectivo centroide. Por

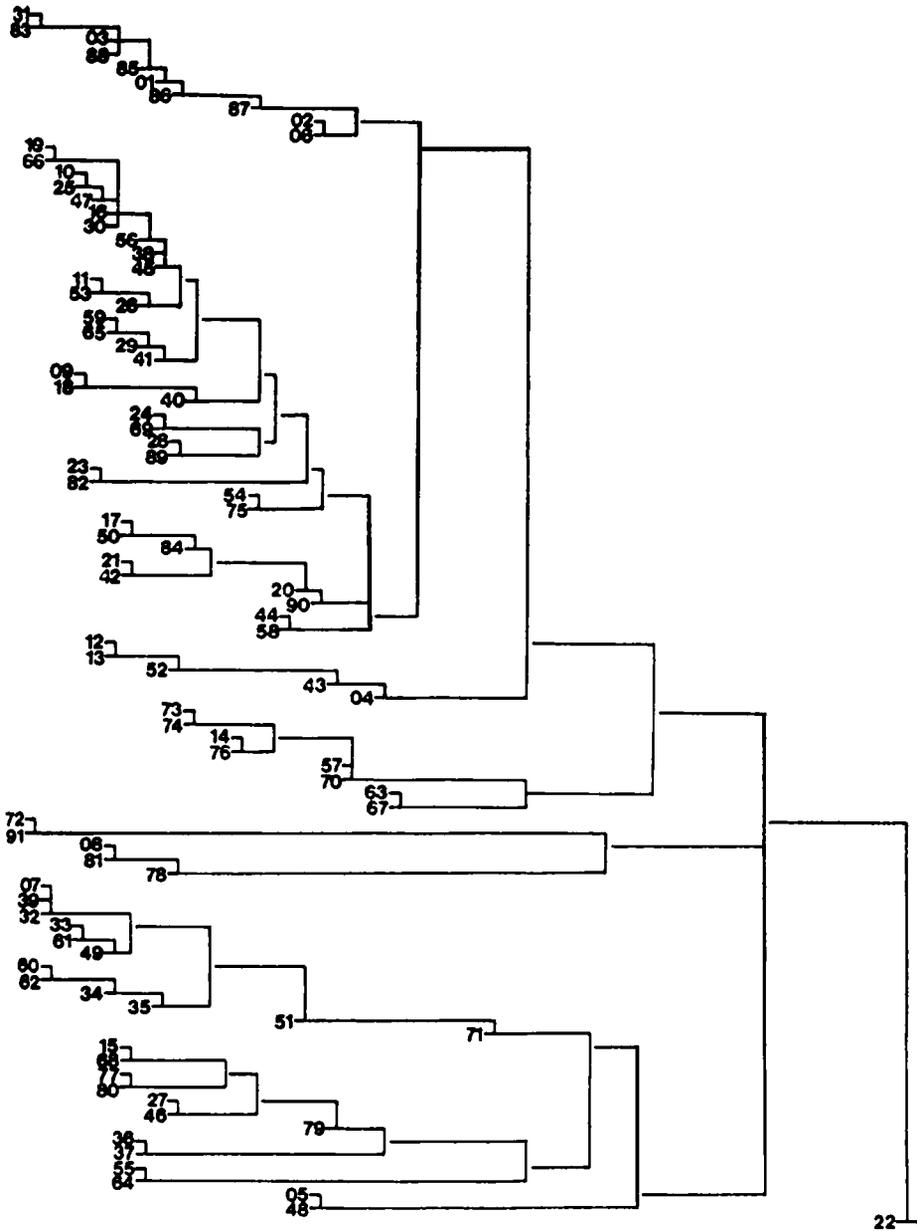


FIG. 5 — Fenograma representativo de 91 amostras de cerâmicas Maia e argilas de Lubaantun, obtido caracterizando-as pelas concentrações de 19 elementos, calculando distâncias euclidianas médias e aplicando o método UPGMA. (Extraído de «Archaeometry» 18, 2 (1976) 147-168).

vezes essa caracterização não é feita através do centroide, mas sim dum constituinte particular de cada grupo (o centrotipo) o qual, se se usarem distâncias euclidianas, é o constituinte mais próximo do centroide.

A distribuição dos OTU's por esses k grupos faz-se cje acordo com um dado critério, como por exemplo o de se reduzir ao mínimo a soma das distâncias de cada constituinte dum grupo ao respectivo centróide (ou centrotipo). Os métodos das k -médias exigem, por isso, o uso de processos iterativos de inspeção, que envolvam a transferência de OTU's dum grupo para outro, progressivamente, de forma a que o critério fixado possa ser obedecido no mais curto intervalo de tempo.

Na maioria dos casos o investigador não conhece a priori qual é o número de grupos mais apropriado para repartir o conjunto inicial de OTU's. Assim, o que se faz normalmente é aplicar o método repetidas vezes, para diferentes valores de k , e observar como varia com k o valor mínimo do critério fixado. O valor de k mais apropriado é então indicado pela mudança pronunciada de inclinação que se verifica a certa altura na curva representativa dessa variação.

8.5 —ANÁLISE EM COMPONENTES PRINCIPAIS

A representação do conjunto de l OTU's num espaço com mais de três dimensões não se pode fazer recorrendo a meios convencionais. Conhecem-se, todavia, alguns métodos que permitem ordenar esse conjunto num espaço a três dimensões, ou mesmo num espaço a duas dimensões, sem distorções apreciáveis.

Um deles é o método de análise em componentes principais, o qual consiste, muito resumidamente, em substituir o conjunto original de atributos por um novo conjunto de atributos não correlacionados (as componentes principais), cujos estados representam uma combinação linear dos estados dos atributos do primeiro conjunto.

Embora este método seja, essencialmente, um método algébrico, é mais facilmente perceptível em termos geométricos, tomando por base as noções de espaço multidimensional e distância atrás referidas.

Suponhamos os OTU's, considerados, representados por pontos num espaço n -dimensional, usando como referencial um sistema de n eixos coordenados com a origem no centro de gravidade do conjunto desses pontos. O método de análise em componentes principais consiste em substituir esse referencial por outro, isto é, por um novo sistema de eixos coordenados ortogonais, com a origem no mesmo centro mas orientados duma maneira mais vantajosa, e em determinar as coordenadas dos pontos representativos desses OTU's em relação ao novo referencial. As direcções escolhidas para os novos eixos são as seguintes: a do primeiro eixo (primeira componente principal) é a direcção segundo a qual se verifica a maior dispersão dos pontos do conjunto; a do segundo eixo (segunda componente principal) é a direcção perpendicular à primeira e a que corresponde a maior dispersão dos mesmos pontos; etc..

Um dos aspectos importantes do método de análise em componentes principais é o facto, atrás apontado, de a representação dos OTU's no espaço das três primeiras componentes, ou mesmo no espaço das duas primeiras, não diferir apreciavelmente, em grande parte dos casos, da sua representação no espaço n -dimensional. Aliás pode medir-se o grau de distorção dessas representações, calculando, por exemplo, a fracção da variância total que lhes corresponde. Assim, torna-se possível visualizar a ordenação dos OTU's no espaço que nos é familiar, sem erros muito grandes. Veja-se, por exemplo, a Fig. 6 onde estão representadas no espaço definido pelas duas primeiras componentes principais e no espaço definido pela primeira e terceira as projecções dos OTU's considerados no estudo a que nos referimos anteriormente, cuja análise de grupos conduziu ao fenograma da Fig. 4. Essa representação tem em conta 69% da variância total, apesar de se ter passado dum espaço a 17 dimensões para outro a 3 dimensões.

Do ponto de vista algébrico o método de análise em componentes principais consiste em calcular:

- a) a matriz R de coeficientes de correlação entre atributos, definida por $R = [1/ (t-1)] XX'$, onde t designa o número de OTU's considerados, X a matriz de dados estandarizados e X' a matriz transposta de X ;

- b) os valores próprios X_1, X_2, \dots, X_r ($r < \min n, (t-1)$) e os correspondentes vectores próprios V_1, V_2, \dots, V_r (componentes principais) da matriz R através da resolução da equação $(R - X I) v = 0$, onde I é a matriz identidade;
- c) a matriz de projecção P , isto é, a matriz das coordenadas dos pontos representativos dos OTU's no novo sistema de eixos, recorrendo à equação $P = Y'X$, onde V' designa a matriz transposta da matriz de vectores próprios convenientemente normalizados.

8.6 — OUTROS MÉTODOS DE ANÁLISE DE DADOS

Deve notar-se que se conhecem muitos outros métodos de análise de dados. Existem, além disso, métodos que permitem testar, do ponto de vista estatístico, a validade dos grupos que se obtêm através da aplicação de métodos de análise de grupos, a que nos referimos em 3.4.

Não iremos, porém, ocupar-nos de mais nenhum, nem isso faria sentido tendo em atenção o objectivo que nos propusemos inicialmente. Na verdade, salvo no que respeita a um deles, ao método de análise em coordenadas principais, que pode em certa medida considerar-se uma variante do método de análise em componentes principais, não estamos por enquanto em condições de os utilizar por falta de programas para o computador.

Ao leitor interessado em conhecê-los e em avaliar as suas possibilidades recomendamos a consulta do livro de Doran e ífodson¹⁰ e das actas dos simpósios referidos na bibliografia^{6,7}.

3.7 _ PROGRAMAS PARA COMPUTADOR

O LFEN possui um programa de análise de grupos, denominado DISSIM²⁶, o qual, por ter sido escrito para o computador PDP 15, com uma memória de apenas 24 K, só pode ser aplicado para classificar 60 OTU's, usando 30 atributos.

Outros programas que estamos a empregar em análise de dados são os seguintes: o sistema de programas MINT²⁷, que é o mais completo a que podemos recorrer neste momento; o programa

NADIST²⁸, que apenas permite calcular matrizes de semelhanças (ou dissemelhanças); e o programa AGCLUS²⁹, que só permite efectuar análise de grupos. O sistema MINT está em condições de correr completamente no computador do Instituto Gulbenkian de Ciência; os dois últimos programas apenas podem correr, por enquanto, no computador do Centro de Cálculo das Universidades de Lisboa. Há que notar que estes dois programas contêm alguns métodos de tratamento de dados que não são considerados no sistema MINT.

Vai procurar-se efectuar, muito brevemente, a adaptação do programa ADCORR³⁰ aos computadores daqueles centros de Cálculo, o qual permitirá efectuar testes estatísticos da validade dos grupos obtidos através da aplicação dos programas de análise de grupos.

3.8 — CONSIDERAÇÕES FINAIS

É importante reconhecer que o agrupamento de artefactos arqueológicos realizado com base em resultados de análises químicas traduz, essencialmente, relações de semelhança química entre esses artefactos.

Crê-se, todavia, que essas relações possam eventualmente pôr em evidência relações significativas do ponto de vista arqueológico, isto é, que elas possam constituir provas valiosas, quer para apoiar ou rejeitar certas hipóteses de trabalho, quer para confirmar ou invalidar conclusões tiradas através da aplicação doutros métodos de investigação.

4. AGRADECIMENTOS

Várias organizações e pessoas têm contribuído, directa ou indirectamente, para o desenvolvimento deste projecto de investigação em curso. Arqueólogos do Museu Monográfico de Conímbriga, do Museu Nacional de Arqueologia e Etnologia, do Museu de Arqueologia e Etnologia da Junta Distrital de Setúbal e do Instituto de Arqueologia da Faculdade de Letras de Lisboa, definindo os temas de trabalho e seleccionando as cerâmicas para

estudo. A Secretaria de Estado da Cultura, mediante a concessão dum subsídio. A Agência Internacional de Energia Atômica, suportando os encargos relacionados com a visita ao Laboratório de Física e Engenharia Nucleares de dois investigadores: o Dr. G. Harbottle, do Departamento de Química do Laboratório Nacional de Brookhaven, a quem se deve a obtenção dos programas NADIST, AGCLUS e ADCORR; e o Dr. J. Op de Beeck, do Instituto de Ciências Nucleares da Universidade de Ghent, que nos ofereceu os seus programas GELIAN, OLIVE e DISSIM. O Instituto Gulbenkian de Ciência, cujo Centro de Cálculo nos tem facilitado a utilização do sistema de programas MINT. O Centro de Cálculo das Universidades de Lisboa, que nos proporcionou a adaptação dos programas NADIST e AGCLUS ao seu computador. A Eng. M. A. Gouveia, do LFEN, que tem colaborado na realização dos ensaios de análise por activação. O Dr. M. B. Lima, da Estação Agronómica Nacional, cujos conselhos e entusiasmo me têm ajudado grandemente no estudo da taxonomia numérica. Aqui lhes expresso os meus agradecimentos.

Agradeço, por fim, ao Prof. R. R. Sokal a gentileza de se ter prontificado a ler a alínea 3 do original e algumas sugestões que contribuíram para tornar mais clara a exposição dessa alínea.

J. M. PEIXOTO CABRAL

5. BIBLIOGRAFIA

- 1 - LIBBY, W. F. (1946). *Atmospheric helium three and radio-carbon from cosmic radiation*. «Phys. Rev.», 69, 671-2.
- 2 - BRAINERD, G. W. (1951). *The place of chronological ordering in archaeological analysis*. «Am. Antiquity», 16, 301-313.
- 3 — ROBINSON, W. S. (1951). *A method for chronologically ordering archaeological deposits*. «Am. Antiquity», 16, 293-301.
- 4 - GARDIN, J. C., ed. (1970). *Archéologie et Calculateurs*
Paris: CNRS.

- 5 — BRILL, R. H., ed. (1971). *Science and Archaeology*. Cambridge, Massachusetts: The MIT Press.
- 6—HODSON, F. R., KENDALL, D. G. & TAUTU, P., eds. (1971). *Mathematics in the Archaeological and Historical Sciences*. Edinburgh: Edinburgh University Press.
- 7 — BORILLO, M., ed. (1972). *Les Méthodes Mathématiques de l'Archéologie*. Marseilles. (CADA), CNRS.
- 8 — BECK, C. W., ed. (1974). *Archaeological Chemistry*. Washington, D. C.: American Chemical Society.
- 9 — AITKEN, M. J. (1974). *Physics and Archaeology*, 2nd ed. Oxford: Clarendon Press.
- 10 — DORAN, J. E. & HODSON, F. R. (1975). *Mathematics and Computers in Archaeology*. Edinburgh: Edinburgh University Press.
- 11 — CABRAL, J. M. P. (1974). *Aplicação do método de análise por activação com neutrões na determinação da proveniência de objectos com interesse arqueológico*. «Arqueologia e História», 5, 361-95.
- 12 — FLANAGAN, F. J. (1973). *1972 values for international geochemical reference samples*. «Geochim. Cosmochim. Acta», 37, 1189-1200.
- 13 — ABASCAL-M., R., HARBOTTLE, G. & SAYRE, E. V. (1974). *Correlation between terra cotta figurines and pottery from the valley of Mexico and source clays by activation analysis*. *Archaeological Chemistry* (ed. C. W. Beck), 81-99. Washington, D. C.: American Chemical Society.
- 14 — OP DE BEECK, J. (1972). *GELIAN program*. Institute for Nuclear Sciences, University of Ghent.
- 15 — OP DE BEECK, J. (1974). *OLIVE program*. Institute for Nuclear Sciences, University of Ghent.
- 16 — SNEATH, P. H. A. & SOKAL, R. R. (1973). *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. San Francisco: W. H. Freeman and Company.

- 17 — SOKAL, R. R. (1961). *Distance as a measure of taxonomic similarity*. «Systematic Zool.», 10, 70-79.
- 18 — GOWER, J. C. (1971). *A general coefficient of similarity and some of its properties*. «Biometrics», 27, 857-871.
- 19 — HARBOTTLE, G. (1976). *Activation analysis in archaeology. A Specialist Periodical Report, Radiochemistry, Vol. 3, 33-72*. London: The Chemical Society.
- 20 — PEARSON, K. (1926). *On the coefficient of racial likeness*. «Biometrika», 18, 105-117.
- 21 — ROHLF, F. J. & SOKAL, R. R. (1965). *Coefficients of correlation and distance in numerical taxonomy*. «Univ. Kansas Sci. Bull.», 45, 3-27.
- 22 — CABRAL, J. M. P. (1977). *Effect of the scaling method on the classification of archaeological ceramics by numerical taxonomy*. Em via de publicação.
- 23 — BIEBER Jr., A. M., BROOKS, D. W., HARBOTTLE, G. & SAYRE, E. V. (1976). *Application of multivariate techniques to analytical data on Aegean ceramics*. «Archaeometry», 18, 59-74.
- 24 — HAMMOND, N., HARBOTTLE, G. & GAZARD, T. (1976). *Neutron activation and statistical analysis of Maya ceramics and clays from Lubaantun, Belize*. «Archaeometry», 18, 147-168.
- 25 — SOKAL, R. R. & ROHLF, F. J. (1962). *The comparison of dendrograms by objective methods*. «Taxon», 11, 33-40.
- 26 — OP DE BEECK, J. (1973). *DISSIM program*. Institute for Nuclear Sciences, University of Ghent.
- 27 — ROHLF, F. J. (1971). *MINT, User's manual*, Dept. of Ecology and Evolution, State University of New York, Stony Brook.
- 28 — BIEBER, Jr., A. M. (1973). *NADIST Program*. Brookhaven National Laboratory, Upton, N. Y.
- 29 — OLIVER, D. C. (1973). *AGCLUS, Aggregative hierarchical clustering program*. Dept. of Psychology and Social Relations, Harvard University, Cambridge, Massachusetts.
- 30 — SAYRE, E. V. (1973). *ADCORR program*. Brookhaven National Laboratory, Upton, N. Y.