Traces et fonction : les gestes retrouvés
Colloque international de Liège
Éditions ERAUL, vol. 50, 1993

Supervised pattern recognition and experimental archaeology

*Amilcare BIETTI**

RÉSUMÉ

Les procédures de classification supervisées en reconnaissance de schémas se caractérisent par des règles de décision *a posteriori*, employant des méthodes statistiques standards, comme, par exemple, l'analyse discriminante de Fisher, la méthode de la vraisemblance maximale, etc. Le but de cette classification est la reconnaissance d'un ensemble « inconnu » d'objets, qu'on appelle le *test set*, à partir d'un ensemble « connu », le *learning set*. Un exemple de ce type de classification est, selon moi, celui de comparaison quantitative entre les données archéologiques et les données expérimentales : c'est le cas des traces d'usure sur les outils archéologiques, comparées avec les traces que l'on obtient par expérimentation sur des outils préparés *ad hoc*. Cet article présente les caractères principaux de la classification supervisée, et ensuite une illustration pratique de la comparaison de deux chaînes opératoires, l'une archéologique et l'autre expérimentale, d'un ensemble du Moustérien italien, en employant un système interactif de reconnaissance de schéma.

ABSTRACT

Supervised classification procedures in pattern recognition consist in the *a posteriori* classification by standard statistical techniques, such as Fisher's discriminant analysis, linear maximum likelihood, etc., of an « unknown » set of objects, the so-called *test* set, in terms of a « known » set of objects, the *learning* set. A classical example of practical application of such procedures may be provided, in my opinion, by the comparison between the archaeological data and the data that have been obtained by experiment, such as those of microwear analysis on archaeological artifacts, and the data obtained experimentally by direct use of *ad hoc* prepared artifacts. In this paper I will shortly discuss the main characteristics of this type of approach, using an interactive system of pattern recognition, and also showing some practical results concerning experimental and archaeological reduction sequences of Mousterian artifacts.

Traditionally, pattern recognition can be defined as the search for structures and statistically meaningful regularities of a set of objects in the feature space (see, for instance, Duda and Hart, 1973) : every object is described in terms of these « features », *i. e.* numerical parameters, both continuous and discrete. These structures, corresponding to complex correlations between features, are often the starting point for the formulation of research hypotheses.

Pattern recognition has been widely used in different fields – biology, biomedics, engeneering, etc. –, as can be seen from the various symposia devoted to the topics (see, for instance, Gelsema, Kanal, 1980, 1986), but only rarely has been applied, in its full sense, to archaeology (see, for instance, Bietti *et al.*, 1985).

In archaeology (and in particular in prehistoric archaeology), pattern recognition is mainly used as a classification of objects. There are, however, two different types of classification procedures : the first one, most widely used in archaeological applications, can be called unsupervised classification, and deals mainly with tentative classifications of data sets where one lacks *a priori* knowledge, where as the second type, supervised classification, is aimed at the recognition of unknown sets of objects, starting from an *a priori* knowledge, the so-called « learning » set. We are thus dealing with an *a posteriori* classification.

The first type of classification is well known in archaeology, in the realm of multivariate procedures, such as cluster analysis, for instance, while supervised classification is practically unknown, up to now (for a first attempt see, for instance, Bietti *et al.*, 1985 : 221-223).

In the next section I will briefly summarize the main capabilities of the supervised classification, using as a reference an interactive system (ISPAHAN) developed for biomedical research (Gelsema,

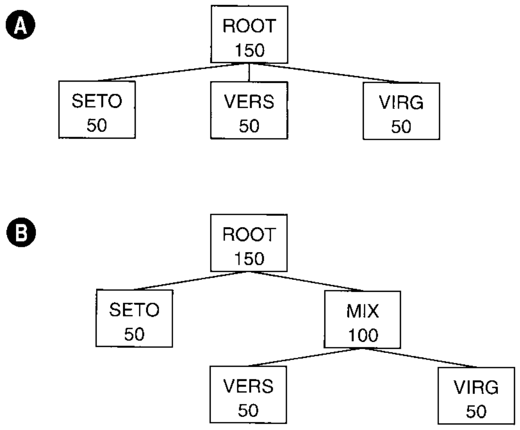


Fig. 1. A. Simple tree structure of the iris set, with the three *a priori* classes : setosa, versicolor and virginica. **B.** A more « developed » tree structure.

1980,1988), and in the third section I will emphasize the importance of supervised classification in experimental archaeology, *i. e.* for the applications where the archaeological data are compared with purposefully « reconstructed » data.

In the same section I will also briefly illustrate an application of supervised classification to the study of the reduction sequence in a Mousterian assemblage of southern Latium (Grotta Breuil), where the main scope of the analysis is the comparison between an archaeological sample of flakes and a similar experimental sample.

The main properties of supervised classification

For a practical illustration of the properties of pattern classification we can take the classical « theoretical » example of Fisher's data set of 150 iris flowers (Agrawala, 1977) subdivided into three classes : setosa, versicolor and virginica. Every flower is characterized by four features, *i. e.* the length and width of the sepal leaves and petal leaves.

The first step is an *a priori* « tentative » classification, and the data may be arranged according to the simple tree structure shown in figure 1 : A, or according to the more « developed » tree structure shown in figure 1 : B, where the versicolor and virginica are considered as « subclasses » of a MIX class.

In principle these two structures are « tentative », *i. e.* they must by checked by statistical analyses that should quantitatively establish if the classes shown in figure 1 are really distinct and meaningful. This is the principal aim of the unsupervised classification. Actually, this term is usually employed for clustering procedures, *i. e.* the *a priori* classification of objects in a multidimensional space (the feature space : in the case of the iris set it is a four-dimensional space). In this paper, however, unsupervised classification will be considered in a broader sense, *i. e.* all the bulk of possible techniques that can be used in order to obtain the best partition into classes (if this is possible) of the data set.

Besides the traditional clustering techniques between the objects, where the features are considered fixed and well defined (Q-mode analysis :

the ISPAHAN system offers a series of well known procedures, such as K-means clustering, nearest neighbor and hierarchical clustering (Gelsema, 1988)), we need also a « screening » of the features, *i. e.* we have to establish which features are diagnostic for the classification : it may well be that some features are irrelevant or are « redundant ».

This may be achieved by standard univariate tests (such as, for instance, the Student's T-test), by multiple partial correlation methods (linear stepwise discriminant analysis) and by R-mode multivariate techniques, such as the well known principal component analysis. We may need, in some cases, where discrete features are also present (this is not the case for the iris set, but, as we shall see, it is often the case in the archaeological data sets), correspondence analysis as well, which is traditionally defined as a Q + R mode multivariate procedure.

These statistical analyses should provide us, in this hypothetical example, with the best partition of objects, *i. e.* a structure of the type illustrated in figure 1 : A or B, and where, of course, as a result of the analyses, the objects are no more necessarily subdivided in equal amounts (as is the case of fig. 1), but, for instance, we have now 48 flowers in the class SET, 56 in the class VERS and 46 in the class VIRG. The resulting tree structure is now a well established structure, that may be called a learning set.

At this point, going on with this « gedanken » experiment, let us suppose that I will grow in my garden a series of iris flowers, say 120, and that I want to know how I can classify these flowers. First of all I will measure the lengths and widths of the sepal and petal leaves, and then draw a tentative *a priori* classification, similar, for instance, to the one illustrated in figure 1 : A. Subsequently, as for the Fisher's set, all the necessary statistical tests will be performed, and we will end up, for instance, with 42 flowers in the class SET, 38 in the class VERS and 40 in the class VIRG.

I want now to establish if this is a correct and significant classification as regards the « reference » data set, *i. e.* the Fisher's set . The set of my flowers becomes a test set, a set on which one has to make *a posteriori* decisions starting from the Fisher's learning set. Such *a posteriori* classification is just what is usually called supervised classification.

There is a series of procedures for supervised classification. The ISPAHAN system, for instance, provides us with five different decision functions :

1. Fisher linear discriminant ;
2. Linear maximum likelihood, minimizing the Mahalanobis distance between the objects of the sets ;
3. Quadratic maximum likelihood ;
4. Bayes, *i. e.* classification according to the rules of Bayes theorem : one has to indicate *a priori* probabilities for each class ;
5. Classification according to the nearest neighbor method.

These different decision functions assume different hypotheses about the distributions of the features : in the case of our iris example this may be irrelevant, whereas, as we shall see in the next section, for the archaeological problems these hypotheses may lead to serious constraints in the application of the classifiers.

In any case, the application of one of the previous decision functions results in a « confusion matrix » which indicates the degree of classification of the set to be tested in terms of the learning set. For our fictitious example of the iris flowers, figure 2 gives the confusion matrix for the flowers of the garden, in the case of the tree structure of figure 1 : A : the classifier has been applied at the node ROOT. As one can see, the classification is perfect for the setosa class, whereas for the virginica and versicolor there is some degree (less then 10 %) of misclassification.

It is worth stressing that, if possible, more supervised decision procedures should be applied to the same learning and test sets : if the confusion matrixes are essentially independent from the decision function that has been employed, we can be reasonably confident in the result of the classification.

As we shall see in the next section, however, in the archaeological problems the use of several decision functions is seriously limited by the nature and the distribution of the features.

Applications to archaeology : archaeological vs. experimental data. A practical example

Unsupervised pattern recognition, in the aforementioned broader sense of tentative *a priori* classification, has been extensively applied in archaeological problems : there are a large number of applications of clustering techniques, and any proceedings of conferences on mathematical and statistical methods in archaeology can provide several excellent examples.

Less common is the application of supervised classification procedures, the reason being, in my opinion, in the lack of the need of *a posteriori* classification in Archaeology : after all, as it is well known, the archaeological data are all derived by unique experiments (the excavation of a site, for instance), so that the idea of a learning set, as it was explained in the previous section, is hardly conceivable.

The situation changes, in my opinion, in the case of experimental archaeology, *i. e.* where the data coming from the archaeological sites have to be compared with data experimentally reproduced.

A classic example is given by the problems arising from microwear analysis, the main topic of the present symposium : experimentation of lithic tools on several materials (bone, wood, hide, etc.) is absolutely necessary in order to have a reference basis of comparison : it is thus clear that, in a quantitative sense, the experiments provide us with a learning set, whereas the archaeological data consitute the test set, according to the definitions given in the last section.

Another example is provided by the technological studies on the lithic industries, and, in particular, the studies of the reduction sequences (« chaînes opératoires ») of the debitage : the experiments, the tentative reproduction of the hypothesized reduction sequences, are ot paramount importance in the understanding of the debitage. Once again, the data obtained by the experiments can be considered as a learning set, and the archaeological data, the test set, may then be classified according to supervised procedures.

Confusion Matrix (Horizontal : posteriori Vertical : priori)			
	Setosa	Versicolor	Virginica
Setosa	42	0	0
Versicolor	0	35	3
Virginica	0	3	37

Fig. 2. Confusion matrix for an hypothetical set of 120 iris flowers, where the learning set is given by the Fisher's set represented in fig. 1-A.

For practical applications the main difficulty comes, in my opinion, from the particular nature of the archaeological data and, more precisely, from the nature of the features involved : such a difficulty is obviously present also in any problem of tentative *a priori* classification. The features defining an archaeological object are, in fact, sometimes continuous (for instance length, width, etc.) and more often discrete (*e. g.* presence/absence of some character), and it is very difficult to make hypotheses about the statistical distribution of these features.

Taking the microwear analysis as an example, any flint specimen is characterized by a set of features regarding the polish. Some of them are continuous (the distribution over the surface, for instance) and many others discrete : the degree of development, the contrast with the unused part of the object, the brightness and the texture, for instance. There are, furthermore, other features connected with the morphology and the longitudinal cross section of the edge, etc.

A priori classification procedures such as Student's T-test (univariate) and principal component analysis (multivariate) are not applicable in this case. The use of stepwise linear discriminant analysis can however provide interesting results, and correspondence analysis is probably one of the best multivariate procedures that can be employed in a mixed situation, with the simultaneous presence of continuous and discrete features.

I will turn now to a factual example. An extended program of research is in progress for the Mousterian site of Grotta Breuil at Monte Circeo, in Southern Latium : besides the studies on stratigraphy, site formation processes, fauna (mostly of archaeozoological character), there is a special branch of research concerning the microwear traces and the technological studies on the reduction sequences in the lithic industry. The microwear studies are carried on by C. Lemorini and have already provided some interesting results (see, for instance, Lemorini, 1990-1991). The total number of artifacts showing microwear traces that has been examined so far is, however, still too limited to represent a suitable basis for a statistical analysis, and more precisely, for reliable applications of supervised classification.

On the other hand, the technological studies on the possible « chaînes opératoires » in the lithic industry can provide us with a sounder statististical

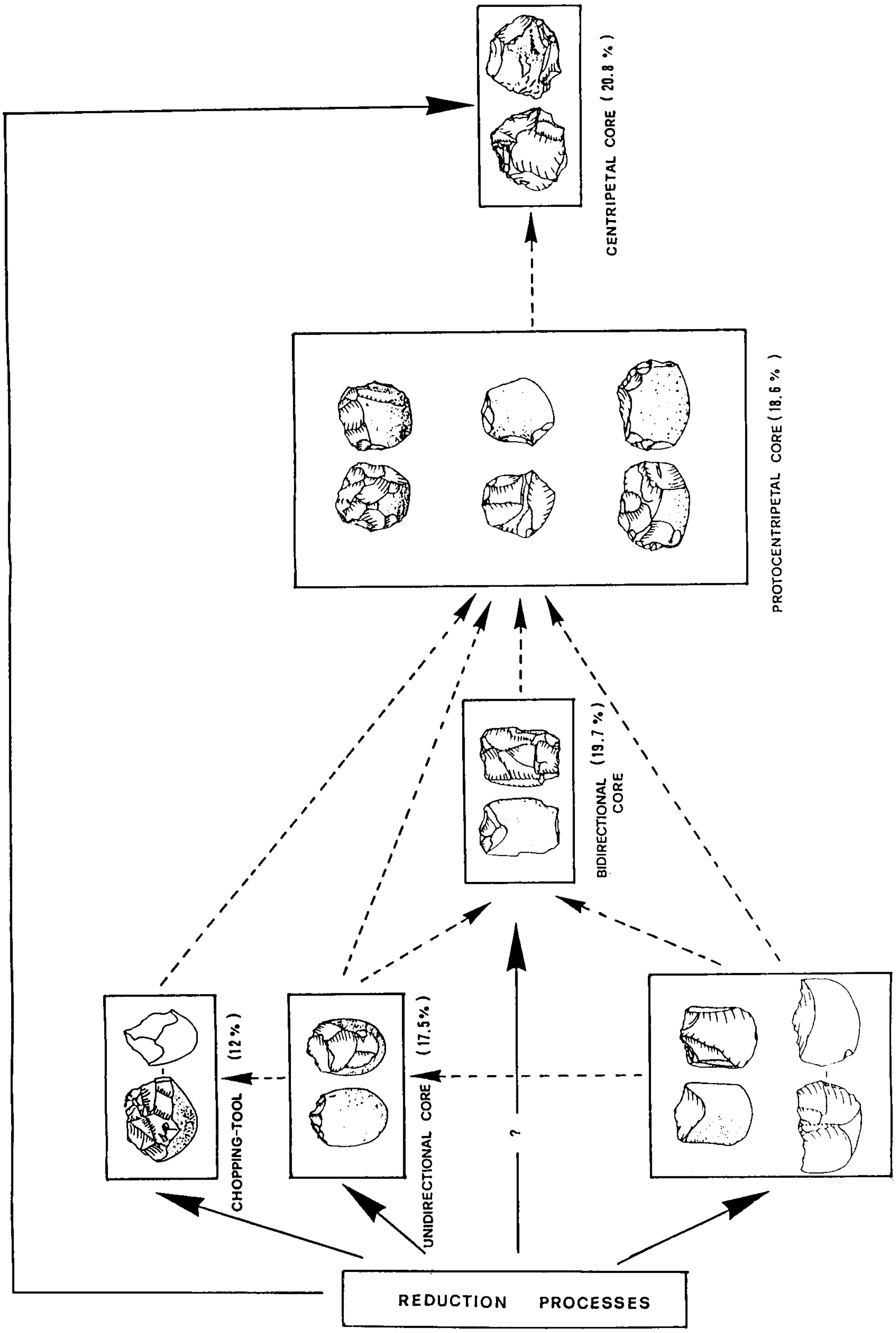


Fig. 3. Tentative reduction sequence for the cores of the lower layer (XX) of Grotta Breuil. (from Rossetti, Zanzi, 1990-1991).

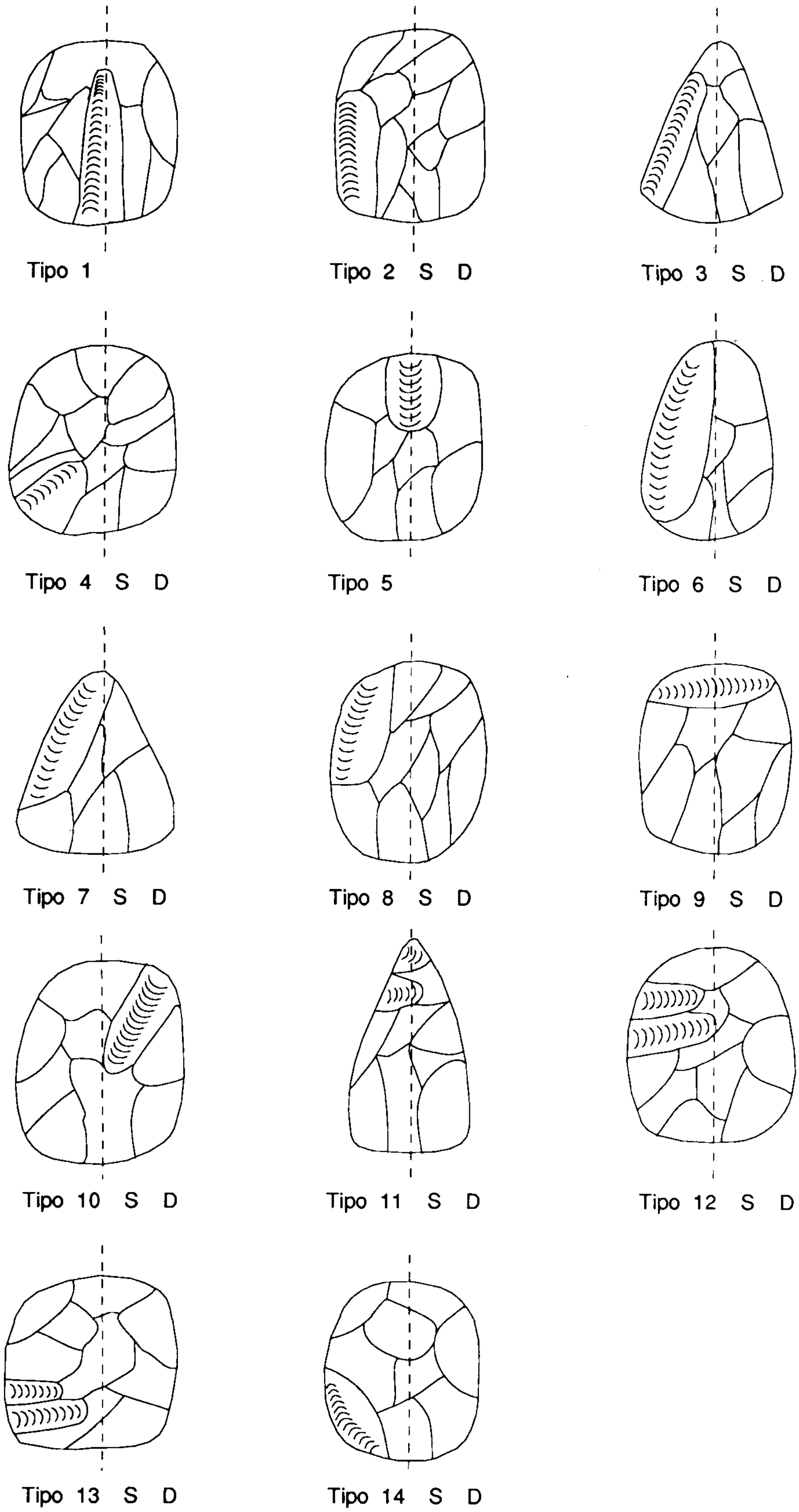


Fig. 4. Schematic classification of the dorsal scar patterns of a flake (adapted from É. Boëda, 1986).

basis, and I will try to illustrate briefly an example of supervised classification derived from the analysis of the attribution of flakes of an archaeological sample, compared with experimental samples.

The archaeological sample has been collected in the lower strata of Grotta Breuil (layer XX, see, for further details, Bietti *et al.*, 1990-1991), and the starting point of the analysis is a tentative reduction sequence of the cores, schematically shown in figure 3 (Rossetti, Zanzi, 1990-1991). The full lines in figure 3 indicate the « primary » and « independent » debitage processes, whereas the dashed lines may be interpreted as « continuous » processes : in this sense, the « protocentripetal » cores could be interpreted as « transition » forms between the unidirectional and bidirectional cores on one side and the centripetal cores as final step of the reduction sequence. For more details on these hypothesized reduction sequences for the industry of Grotta Breuil see, for instance (Rossetti, Zanzi, 1990-1991 ; Bietti *et al.*, 1991).

Our problem is the attribution of the flakes to the hypothized chain shown in figure 3 : in other words, the study of the flakes should give us information which is complementary to that obtained from the analysis of the cores.

The useful parameters, the features, are defined not only by the standard metric variables of the flakes, such as length, width and thickness, but also from a series of discrete variables, such as the type of butt and, in particular, the *patterns of dorsal scars*.

The starting point for the classification of the dorsal scars is the work on the Levallois debitage done by É. Boëda (1986), and in figure 4 one can see a schematic classification of the different types of dorsal scars. Ten discrete features have been defined starting from the « template » shown in figure 4 considering longitudinal, lateral and oblique (both left and right) scars. For more details on the definition of these features, see Bietti, Grimaldi, 1990-1991).

The archaeological sample consists of 132 « recognizable » flakes, and in figure 5 : A one can see a tentative tree-structure (the analogue of the iris set in fig. 1 : A). *A priori* classification procedures have been performed on this set of data and in figure 6 one can see the scatter plot along the first two principal axes of the correspondence analysis : the flakes *a priori* attributed to the unidirectional and bidirectional cores are rather well separated from the ones attributed to the protocentriptal (cores with crossed striking platforms) and centripetal cores.

In figure 5 : B one can see the analogous tree structure for an experimental set of 87 flakes : the experiment was performed by S. Grimaldi in the spirit of the continuous lines of figure 3, *i. e.* for a series of « primary » and « independent » reduction processes. Correspondence analysis was also applied to this experimental set and the results were similar to the ones obtained for the archaeological set shown in figure 6.

We can now go further and try to perform a supervised classification : the learning set is given by the experimental structure (fig. 5 : B) and the test set by the archaeological one (fig. 5 : A). Before applying the supervised decision functions we can also do some « screening » of the features ; this may be easily done by means of the linear stepwise discriminant analysis : the result is that out of 14 features 9 are diagnostic for the two sets

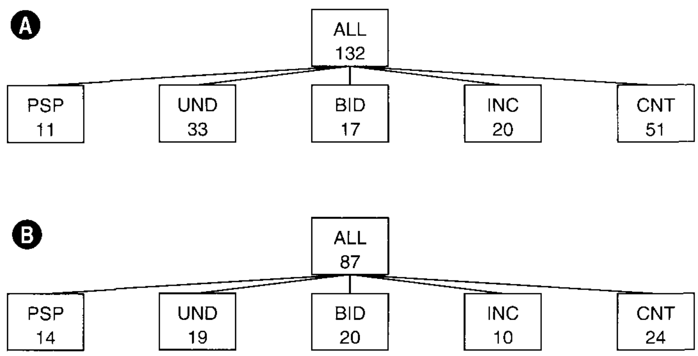


Fig. 5. A. tree structure for a sample of 132 flakes from the layer XX of Grotta Breuil, attributed to different types of cores. PSP : pseudosprismatic ; UND : unidirectional ; BID : bidirectional ; INC : crossed platforms (protocentripetal) ; CNT : centripetal. **B.** Same as A for an experimental sample of 87 flakes.

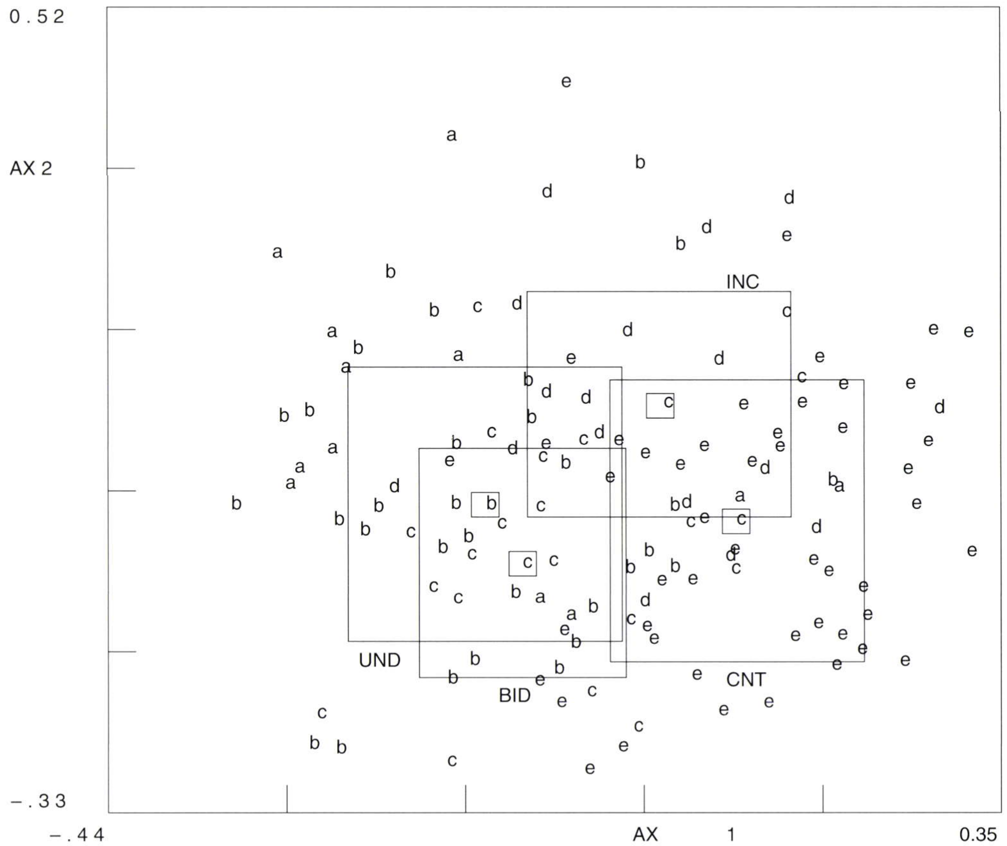


Fig. 6. Scatterplot of the flakes of the archaeological sample according to the first and second principal axes of correspondence analysis : the large rectangles represent ± one standard deviation around the mean (small rectangles) (from Bietti and Grimaldi, 1990-1991).

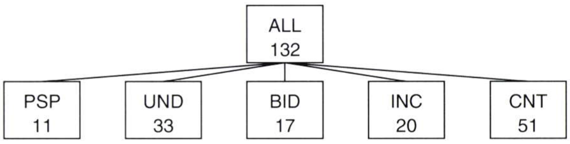
and 6 are in common (for more details see Bietti and Grimaldi, 1990-1991).

The choice of the supervised decision function to be applied is unfortunately rather limited, and it depends, as we have previously stressed, on the nature of the archaeological features. Fisher discriminant analysis assumes normal distributions for the features, and the same holds for the maximum likelihood methods. Moreover they are best suited for continuous features, as it is also the nearest neighbor classifier.

Bayes' decision rule, on the other hand, assumes only the independence of the features, and, moreover, is well suited for discrete features.

In figure 7 (bottom) one can see the confusion matrix using Bayes' decision rule for the tree structure represented at the top of the same figure. This is obviously the test set : the learning set is the one represented in figure 5 : B. The results are rather good for the flakes attributed to the pseudoprismatic and to the bidirectional cores but much worse for the flakes attributed to the protocentripetal and centripetal cores.

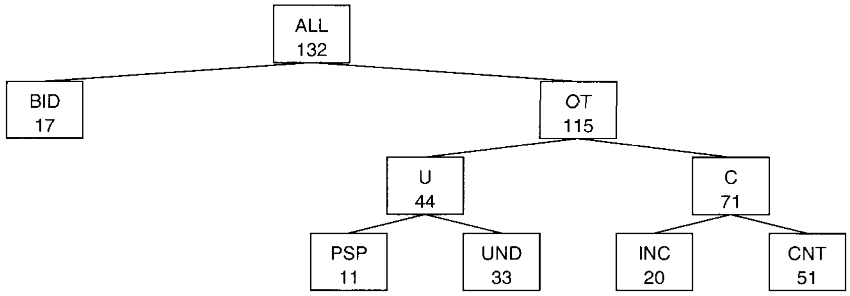
We have also tried to apply Bayes decision rule to another tree structure, which is shown in figure 8 (top). Another experiment, more in the spirit of



Confusion Matrix
(Horizontal : Posteriori Vertical : priori)

	PSP	UND	BID	INC	CNT
PSP	9	1	0	0	1
UND	3	17	6	1	6
BID	0	1	12	2	2
INC	0	10	5	0	5
CNT	0	6	15	5	25

Fig. 7. (top). Tree structure as in fig. 5- A. (bottom) Confusion matrix for the archaeological data using Bayes' decision rule at the node ALL and the structure shown in fig. 5-B as learning set.



A

	U	C
PSP	11	0
UND	33	0
BID	0	0
INC	4	16
CNT	17	34

B

	U	C
PSP	11	0
UND	33	0
BID	0	0
INC	4	16
CNT	17	34

C

	U	C
PSP	11	0
UND	33	0
BID	0	0
INC	4	16
CNT	17	34

D

	U	C
PSP	11	0
UND	33	0
BID	0	0
INC	4	16
CNT	17	34

Fig. 8. (top). Developed tree structure for the same archaeological sample illustrated in fig. 5-A. (bottom) Confusion matrixes obtained using Bayes decision rule with. **A.** a second experiment with 82 flakes as learning set at the node OT. **B.** The first experiment (the set shown in fig. 5-B) as learning set at the node OT. **C.** Same as *a* at the node C. **D.** Same as B at the node C. For further explanations see text.

the dashed lines of figure 3, *i. e.* of a « continuous » reduction sequence, consisiting of 82 flakes, has been performed : we have thus two learning sets, one for each experiment, with the same structure of the test set shown in figure 8 (top). Bayes' decision rule has been now applied to the nodes OT and C : the flakes derived from bidirectional cores have been excluded in the analysis because the second experiment does not really consider a « transition » through these types of cores.

In figure 8 : A one can see the confusion matrix with the second experiment as a learning set at the node OT : the classification is perfect for the UND and PSP flakes, rather good for the INC and fair (33 % of misclassification) for the CNT. The results at the same node with the first experiment as a learning set are shown in figure 8 : B : they are still good for the UND and PSP flakes, but the degree of misclassification for the INC and CNT is reversed in comparison with figure 8 : A. In figure 8 : C and D we have the analogous of figure 8 : A and B when the Bayes' rule is applied to the node C : the degree of misclassification of the CNT flakes is always the same for the two experiments (32 % and 39 %) but the second experiment is more successful for the classification of the INC flakes.

As a conclusion, the two experiments provide essentially similar results, so that we are not yet able to decide, from the analysis of the flakes, if the reduction sequence is more « continuous » or more « independent ». In any case, the supervised classification, especially the one shown in figure 8, is in agreement with the *a priori* result given by correspondence analysis (fig. 6) : it would be very surprising if the *a posteriori* classification resulted in strong disagreement with the *a priori* one !

As I have already stressed, the nature of the data establishes serious limits on the application of several decision functions. I can only add that the application of the linear maximum likelihood has given results somewhat worse than the ones obtained by means of the Bayes' rule, but still acceptable, for the tree structure shown in figure 8, for the UND and PSP flakes.

I have described this example only as a demonstration of the capabilities and of the usefulness of supervised classification when we are dealing with comparisons between archaeological data and experimental data. The sample chosen is rather peculiar : flakes in a very specialized context, i. e. the Mousterian of Southern Latium, the so-called Pontinian, where the industry is made from small flint pebbles.

I am sure, however, that supervised classification will be able to cover a wide range of archaeological subjects, and I emphasize again its usefulness when we are dealing with the recognition of properties of archaeological data sets starting from the study of the properties of experimental sets.

* *Sez. di Antropologia, Dipartimento di Biologia Animale e dell'Uomo, Università di Roma « La Sapienza », P. A. Moro 5, 00185 Roma, Italy.*

Bibliography

AGRAWALA (A. K.), (Ed.), 1977.– *Machine Recognition of Patterns.* IEEE press.

BIETTI (A.), BURANI (A.), ZANELLO (L.), 1985.– *Interactive pattern recognition in Prehistoric Archaeology : some applications. In* : A. Voorrips, S. Loving (Ed.), *To Pattern the Past,* Proc. of the Symposium of the IV Commission UISPP « PACT » series 11, p. 205-228.

BIETTI (A.), GRIMALDI (S.), 1990-1991.– *Patterns of reduction sequences at Grotta Breuil : statistical analyses and comparisons of archaeological vs. experimental data. In* : Proc. of the Intern. symposium. *The fossil man of Monte Circeo : fifty years of studies on the Neandertals in Latium.* Sabaudia, Oct. 1989, Quaternaria Nova I, p. 379-406.

BIETTI (A.), GRIMALDI (S.), MANCINI (V.), ROSSETTI (P.), ZANZI (G.), 1991.– Chaînes opératoires et expérimentation : quelques exemples du Moustérien de l'Italie centrale. XI[e] Rencontres Internationales d'Archéologie et d'Histoire d'Antibes, *In* : *25 ans d'études technologiques en Préhistoire,* Éd. APDCA. Juan-les-Pins, oct. 1990, p. 109-124.

BIETTI (A.), KUHN (S.), SEGRE (A. G.), STINER (M.), 1990-1991.– *Grotta Breuil : general introduction and stratigraphy. In* : Proc. of the Intern. symposium. *The fossil man of Monte Circeo : fifty years of studies on the Neandertals in Latium.* Sabaudia, Oct. 1989, Quaternaria Nova I, p. 305-323.

BOËDA (É.), 1986.– *Approche technologique du concept Levallois et évaluation de son champ d'application.* Thèse de doctorat, Université de Paris X.

DUDA (R. C.), HART (P. E.), 1973.– *Pattern Classification and Scene Analysis.* J. Wiley, New York.

GELSEMA (E. S.), 1980.– *ISPAHAN, an interactive system for pattern analysis : Structures and capabilities. In* : Gelsema and Kanal (1980).

GELSEMA (E. S.), 1988.– *ISPAHAN User Manual.* Department of Medical Informatics, Erasmus University, Rotterdam.

GELSEMA (E. S.), KANAL (L. N.), (Ed.), 1980.– *Pattern Recognition in Practice.* North Holland, Amsterdam.

GELSEMA (E. S.), KANAL (L. N.), (Ed.), 1986.– *Pattern Recognition in Practice II.* Elsevier (North Holland), Amsterdam.

LEMORINI (C.), 1990-1991.– *Prospects for a functional evaluation of the Mousterian site of Grotta Breuil (Monte Circeo, Italy). In* : Proc. of the Intern. symposium. *The fossil man of Monte Circeo : fifty years of studies on the Neandertals in Latium.* Sabaudia, Oct. 1989, Quaternaria Nova I, p. 407-428.

ROSSETTI (P.), ZANZI (G.), 1990-1991.– *Technological approach to reduction sequences of the lithic industry from Grotta Breuil. In* : Proc. of the Intern. symposium. *The fossil man of Monte Circeo : fifty years of studies on the Neandertals in Latium.* Sabaudia, Oct. 1989, Quaternaria Nova I, p. 351-365.