ABSTRACT: Although the discovery procedure for research on unknown symbolic (=semiotic) systems is yet to be found, we present a general method aimed to demonstrate the existence of common structural features in the paintings and engravings of most French and Spanish Palaeolithic caves. We believe that this method is not specific to the European Palaeolithic Cave Art and might be applied to the other rock arts. Our approach should not be related to the decipherment of unknown writing systems. The following scientific fields and methods are discussed: 1. Structural semiotics (semantic oppositions, communication system), 2. Statistical Data Analysis (Factor Analysis and Clustering Algorithms), 3. Natural Language Processing (Language Parsing and Grammar Inference), 4. Knowledge Discovery from databases (Rough Set Approximation), 5. Statistical Comparative Approach (Comparison of palaeolithical rock art with ethnological data from Africa and Australia).

1. Introduction

Using some very simple statistics, Leroi-Gourhan (1965) had attempted to demonstrate that Palaeolithic animal and non-figurative drawings (32,000 to 10,000 years Before Present) were not randomly distributed as regards their associations and locations in caves. We claim that if rock art conveys a collective thought, it should be structured as any other communication system.

2. Structural Semiotics — Typology and Association Rules of Non-Figurative Drawings

TASK: to determine whether non-figurative art is compositional and to establish a typology of all non figurative patterns.

DATA: handmade database of non-figurative drawings.

Instead of considering non figurative (or "abstract") drawings as isolated schematic representations of some everyday use objects, we proposed to consider them as parts of a semiotic system by highlighting some regular relations between them.

Ignoring everything about the semantic content of non figurative drawings we chose to use formal geometric criteria of classification. This method led us to propose 14 keys allowing to classify all Palaeolithic non figurative drawings. Proceeding this way, we had to consider some features as non relevant and even so we obtained about 3% of unclassifiable "abstract" drawings. We consider our classification as one possible hypothesis that should be revised as research progresses. We could observe that most of the proposed types appear in almost all regions. Quadrilateral patterns seem to be universal but some of their features are different in Spain and France. We never found more than four specific features in one quadrilateral pattern. Moreover some of these constitutive features appear sometimes as independent patterns. Thus we could conclude that there existed what we called an articulation of patterns and we distinguished three types of their composition: integration, superposition and juxtaposition. We hypothesized that different types of pattern compositions (signifiers) could be interpreted as referring to different contents (signified). We compared the formal relations in the internal structure of Palaeolithic non figurative drawings with relations found in ideographic writing systems. Doing so, we did not intend to borrow any semantic contents from writing systems but to observe some formal devices (used independently of their contents). At the same time the use of such devices gave evidence that Prehistoric Man was capable of second degree abstractions consisting in establishing relations between abstract patterns.

Palaeolithic rock non-figurative drawings differ considerably from writing systems in that they are clearly non linear. On the
other hand, the great majority of "abstract" patterns appear in panels associated with animal drawings.

There are some arguments which give evidence of that animal motifs and abstract patterns can be considered as belonging to two intertwined semiotic systems. Moreover, negative handprints should also be probably treated as part of one complex semiotic system.

RESULTS: typology, composition types and association rules.

Graphical categories (motifs). The figurative components of European Palaeolithic cave art may be divided into fourteen main motifs. A database (containing 3300 figurative images, coming from 84 caves and rock shelters) consists of 416 polythematic panels (from two to five different themes) was analysed from a statistical point of view. To demonstrate whether these thousands of figures have been put together according to some rules or not, some statistical treatment is necessary.

The thematic composition of our 1027 panels is never very complex 60% of the panels are monothematic (416 panels only contain two themes or more): and the frequency of the different types of panels decreases rapidly as the complexity increases. No panel contains more than six different themes on the fourteen possible. As it is well known in the Theory of Communication, the most simple is the most frequent (Mandelbrot 1954).

Factor Analysis (Benzécrit 1984). According to Factor Analysis, themes that carry the strongest contribution to the definition of axis 1 are hind, stag and ox on the one hand and mammoth, reindeer and bear on the other hand. This axis might be partly explained by regional particularity, hinds and stags being more abundant in Cantabria (Spain) whereas mammoths are mainly abundant in the south-west of France. The second axis separates the mammoth from rare themes: lions, fish and the artificial category “various”. It is noteworthy that the horse, the bison and the ibex are almost at the centre of gravity of the cloud of points.

Hierarchical Cluster Analysis (Jambu 1978). Factor Analysis is usually completed by Cluster Analysis. The tree-structure which is obtained clearly confirms the clustering which was already apparent on the Factor Analysis. Groups (or classes) are established very soon during the analysis. Thus, “cutting” the representation tree at level 2, we get five independent classes:

<table>
<thead>
<tr>
<th>Class</th>
<th>Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>horse, ibex and bison</td>
</tr>
<tr>
<td>2</td>
<td>anthropomorph</td>
</tr>
<tr>
<td>3</td>
<td>mammoth, rhino, reindeer and bear</td>
</tr>
<tr>
<td>4</td>
<td>fish, “various” and lion</td>
</tr>
<tr>
<td>5</td>
<td>stag, hind and ox</td>
</tr>
</tbody>
</table>

Spanning Tree Algorithm (Kruskal 1956). The consistency of these five classes was also shown by applying Kruskal algorithm to the table of co-occurrences of themes. This algorithm is able to extract the "minimum-cost spanning tree" from the complete graph, i.e. the graph in which all the possible links between themes are taken into consideration.

RESULTS: classification of figurative motifs.


TASK: to classify automatically the figurative animal themes in order to determine their interthematic associations.

DATA: database with description of panels reduced to the figurative themes.
of the system (note that the above model produces only 336 combinations out of 6461 possibilities). This shows clearly that the assemblages are governed by rules.

**Context-free Grammar.** The rules expressed in our OR-Graph Grammar above represent the core of the system in an abstract and probabilistic way which says nothing about the constraints. For this reason, we wrote a DCG (which defines a dyad as the sum of two terminals, a triad as a dyad + a terminal etc, the terminals being the 14 figurative motifs).

Thanks to the Cluster Analysis, we could notice that the associations of themes are highly dependent upon the classes they belong to. It is therefore possible to use these classes to formulate “structural constraints”. A very simple DCG grammar consisting of five rewriting rules and a few constraints produces four times less combinations than an unconstrained one. Yet 98% of the combinations attested in the caves are still produced.

**RESULTS:** one constraint-based context-free grammar (using Definite Clause Grammar formalism incorporated in Prolog) and one automatically inferenced OR-Graph grammar.

### 5. Knowledge Discovery in Databases — Rough Set Approximations (2001)

**TASK:** to describe globally all the panels of the figurative art.

**DATA:** relational description of panels with intrinsic and extrinsic features.

Using **Rough Set Approximation algorithms** (Pawlak 1982, 1991), we processed data this time describing figures and consisting of the following attributes: parts of the body (7 values: whole animal, protome, acephalous, without legs, forequaters, hindquaters, back line), size (5 values), orientation (13 values), technique (14 values), animation (5 values). At the present stage of research we do not take into account non figurative motifs which, however, we believe so, constitute a whole system with the figurative motifs.

We found Pawlak’s **Rough Set Theory** particularly fit for our problem: namely it allows to determine within a set of unknown objects which one of their attributes displays the highest discriminating power. As a result of the application of rough set algorithms to our data we ordered the five attributes taken into account in the following way:

- **body part > orientation > size > technique > animation**

This order remains the same for the following most frequent animal (horse, bison and ibex) and it can be considered as a constant of the Palaeolithic figurative system.

Very different species of animals are concerned by the attribute “part of the body”. When limiting our comparison to the two most frequent animals in our database (horse and bison) we could observe that the theme of horse was represented much more often by protome than the theme of bison (the latter were very often acephalous or only their back line was drawn).

Orientation of the animals is the second most important attribute. It appears to play an important role in the composition of panels. As regards all animal species except horse, most figures are left-oriented. But horses appear to be an exception: 55.8% of horses are depicted by their right profile.

**Figure 3.** Factor Analysis of a table of co-occurrences of the 14 figurative motifs.
If we consider the orientation of animals appearing as immediate neighbours in a panel, we can observe that animals belonging to the same species are more frequently oriented in the same direction rather than opposed to each other. On the other hand, when two animals follow each other the bison follows the horse more frequently than the opposite; when bison and horse are opposed to each other they are more frequently crossed than confronted. But our checkbench database is still too small in order that we can consider these observations as certain.

RESULTS: discrimination of a few important descriptive features, applying KDD methods with Rough Set calculi to a significantly enriched testbench information database of Franco-Cantabrian Upper Palaeolithic Art.

6. Perspectives

TASK : to discriminate (regional and diachronical) sub-structures

DATA : further refinement of data (associating figurative with non-figurative drawings)

The next step of this research is to build a huge database in which the individual figures are described by as many attributes as possible. This task is in progress but it is difficult for two reasons. First, some of the necessary information is missing in the literature and must be gathered. Second, in our future research we should utilize more Data Mining Technologies and Machine Learning Algorithms adapting them very carefully to the artistic domain.

On the other hand, we are presently making research on statistical comparison of Palaeolithic rock art with Ethnological data from Africa and Australia.

7. Conclusion

The possibility of a semantic interpretation is the most difficult to achieve in an objective way: we are well aware of the necessity to avoid non documented analogies with any other meaning system (for instance religious beliefs of shamanic or totemic type). Taking into account as many attributes as possible (colour, technique, size etc.) we can give evidence of complex relations (similarity or opposition) between figures as a testimony of human intelligence from more than 30,000 years ago but their semantic content will probably remain unknown for ever.

Let us resume some important issues:

1) the number of figurative motifs in European Palaeolithic Cave Art is small. Some motifs like horse, bison and ibex are true leitmotifs (altogether they represent more than 60% of the imagery);
2) the associations of figurative themes that have been produced represent a very small part of the possible combinations;
3) these features may be accounted for by a formal grammar consisting of a small number of rules. A very simple and coherent model has been derived, which accounts for the thematic composition of 98% of the Palaeolithic Cave Art productions.

The fact that a small number of coherent and recursive rules may account for the whole of the corpus implies that the same rules have been in use during millennia in the very large area featuring Palaeolithic cave art.

Our strategy was to proceed step by step, by successive approximations. As the first approximation, our goal was to answer an important and disputed question concerning the existence of a common structure in Palaeolithic cave art.

The Rough Set theory provides algorithms to determine automatically the most discriminating attributes and to refine progressively the list of useful criteria. Applying this method allowed us to point out the importance of topological and relational attributes such as the orientation of motifs and their relative location. It is interesting to note that the two most important motifs, horse and bison, display with respect to these criteria different properties, which may be interpreted as their compositional features. We are just beginning to discover the basis of a figurative syntax.

References

