24

VANDAL: an expert system for the provenance determination of archaeological ceramics based on INAA data

Vanda Vitali*

Marie-Salome Lagrange†

24.1 Introduction

This paper presents a part of a larger study centering on the role of archaeometric research in the reconstruction of the past.

Archaeological investigations are multidisciplinary endeavors, requiring the input and knowledge of various disciplines at various stages of an investigation. To explore the organization of multidisciplinary archaeological knowledge, particularly the interpretive phase of an archaeological research project, we chose the methodology of constructing an expert system. Studies have shown that expert systems represent very useful tools for rendering the structure of interpretive arguments in archaeological studies explicit and transparent, as well as showing clearly the premises on which archaeological interpretations are based and the way these premises are connected (Lagrange & Renaud 1985, Doran 1986, Gardin *et al.* 1987). This approach thus elucidates the interaction between the different kinds of knowledge involved and determines the kind of knowledge upon which an archaeological interpretation is based.

A phase of this broader investigation consisted of constructing an expert system that would perform an archaeological task involving knowledge in various disciplines. The task selected—the provenance determination of archaeological ceramics based on instrumental neutron activation analysis (INAA)—was chosen for several reasons. In recent years INAA, coupled with multivariate data analysis methods, has become one of the principal tools for provenance determination in archaeology. In order to fully and

* Department of Metallurgy and Materials Science,
University of Toronto,
Toronto
Ontario M5S 1A4

† Centre National de la Recherche Scientifique,
E.R. 315 Paris,
23 rue dur Maroc
75940 Paris Cedex 19
France

properly utilize such a tool, expertise in archaeology, materials science, instrumental analytical chemistry, and data analysis all have to be combined.

From a technical standpoint, archaeometric provenance determination is a relatively structured domain. It is, therefore, thought to be feasible for the development of an expert system. It has also been shown that classification tasks, to which provenance studies belong, represent perhaps the most successful area of application for this tool to date. While it can be argued that such tasks may not be the most sophisticated use of expert systems, our goal in this phase of the research was to produce a functioning system capable of performing a specific archaeological task. It was hoped that the expert system developed, by embodying knowledge from various fields, would prove to be a successful and useful guide through a multidisciplinary archaeological study.

24.2 Approach

The expert system that we developed determines relationships among groups of archaeological ceramics based on INAA and data analysis information, in a way that corresponds to an investigation of Zagros ceramics (Vitali *et al.* 1987). The Zagros archaeological study dealt with the development of a methodology for provenance determination as well as with the specific problems of local production and trade of Chalcolithic ceramics from the central Zagros. It was conceived and carried out by a multidisciplinary group composed of archaeologists, a material scientist, a chemist and a data analyst.

The first stage in assembling the knowledge and structuring the approach to provenance determination involved producing an outline of the reasoning mechanism employed in analyzing information on the chemical composition of ceramics. This was done in a way similar to logicist analysis of archaeological studies (Gardin 1980). Such analysis has been employed in the past for structuring archaeological knowledge prior to a formalization of that knowledge using expert systems (Gardin *et al.* 1987).

The starting information represents a mixture of archaeological, chemical, and data analysis information, and presupposes the correctness of that information. The archaeological information gives the location (site) where the ceramics were found, the time period to which the ceramics were assigned (based on stratigraphic evidence), their category (type), and the relative percentage of each type of ceramic found at each location. The technical information is based on the chemical composition of the ceramics in terms of their elemental concentration for 15 minor and trace elements (Vitali *et al.* 1987). This information was summarized using appropriate data analysis procedures (linear discriminant function and a jackknife procedure for re-classification) and expressed as a misclassification rate.

A comparison of the chemical composition of different archaeological ceramic groups is used throughout the reasoning process to arrive at the different archaeological conclusions given below.

A. Production Method If the chemical composition of any two ceramic groups is indistinguishable, the ceramic groups are considered to have been made by the *same production method* (using the same raw materials and the same technology). If their composition is distinguishable, they are considered to have been made using different production methods.

B. 1. For ceramics of the same production method :

1. If the groups are of the same type, belong to the same site but are from two consecutive time periods, then there is a *continuity of production method* through the specified time periods.
2. If the groups of ceramics belong to the same site and the same period but are different in type, then the *same method* was employed for the production of *different wares;*
3. If the groups of ceramics are of the same type and the same period but are from different sites, then they came from the *same production center.*
   (a) If the groups of ceramics represent only a small percentage of the total ceramic assemblage found at each site, then the ceramics are *imported* to those sites.
   (b) If one of the groups of ceramics represents only a small percentage of the total ceramic assemblage found at one site and the other represents a large percentage of the total ceramic assemblage found at the other site, then the ceramics were *traded* from the site where they were found in a large percentage to the site where they were found in a small percentage.

B. 2. For ceramics of a different production method :

1. If the groups of ceramics are of the same type, belong to the same site but are from two consecutive time periods, then there is a *discontinuity of production method* through the specified time periods.
2. If the groups of ceramics belong to the same site, the same period, but are different in type, then a *different method* was employed for the production of *different wares;*
3. If the groups of ceramics are of the same type, the same period, but from different sites, then they came from *different production centers.*
   (a) If, in addition, a ceramic group represents a large percentage of the overall ceramic assemblage found at the site, then the ceramics were *locally* produced at the site.

Having assembled the required knowledge from domain experts and organized it, the next step in this study involved choosing an appropriate representation formalism and language for building our expert system.

There are several knowledge representation formalisms and languages available for constructing an expert system. The choice may influence the performance of an expert system and thus it has to be made with regard to the domain knowledge and the task to be performed (Reichgelt & van Harmelen 1986, Cerri *et al.* 1987). However, for multidisciplinary domains such as ours, no single formalism or language is likely to provide the optimal solution.

For the construction of our first operational prototype expert system, named VANDAL, we selected the SNARK language and its associated inference engine (Lauriere 1986, Lagrange & Renaud 1987). SNARK has been used in the past for the construction of several prototype expert system in archaeology (Gardin *et al.* 1987).

SNARK language is based on first-order logic. Its syntax for the description of data is **a** R **b** where **a** designates an entity, **R** designates a relationship between **a** and **b**, and **b** designates a value. The semantic content of the formulations is left to the user. The order in which the data is presented is unimportant. The advantages of this type of representation is that it is relatively easy to understand for a user and that it is flexible.

An example of a set of statements describing a group of ceramics in VANDAL may contain the following:

K-SG-HARD-P1	NATURE	WARE
K-SG-HARD-P1	TYPE	HARD
K-SG-HARD-P1	INCLUDE	BLACK ON BUFF
K-SG-HARD-P1	INCLUDE	UNTEMPERED
K-SG-HARD-P1	SITE	SEH GABI
K-SG-HARD-P1	PERIOD	MC1
K-SG-HARD-P1	CONSECPERIOD	MC2
K-SG-HARD-P1	PERCENT	7
MISCLASS-R-1	COMPARE	K-SG-HARD-P1
MISCLASS-R-1	COMPARE	K-GD-HARD-P1
MISCLASS-R-1	VALUE	0.05

The knowledge base in SNARK is represented as a production system or sets of rules. In this representational scheme, causal relationships, which are all binary, are expressed in the form IF x THEN y, where x is a condition and y is an action. The syntax of a *condition* is **SI R(a)*b (if R(a)*b)** where R designates a relationship, **a** designates an entity, **b** designates a value, and * is an operand (such as =, >, <, etc.). The syntax of an *action* is **ALORS R(a) <– b (then R(a) <— b)** where <— is an assignment symbol. While the rules can be written in any order, there is a way to control the order in which they are examined.

It can be argued that the rule-based production systems are the least adequate means of knowledge representation available in expert systems (Doran 1988). However, they can be easily presented in modules, coded, examined and modified by all domain experts and thus appear to be the easiest way to group and organize the knowledge from various experts in the initial stages of building VANDAL expert system.

An example of a VANDAL rule is:

REGLE: SAME PRODUCTION METHOD
(rule:)

SI	NATURE	(A) = WARE
	(&) NATURE	(B) = WARE
	(&) CHEM-COMP	(A) = (X)
	(&) CHEM-COMP	(B) = (Y)
	(&) NATURE	(C) = MISCLASS-RATE
	(&) COMPARE	(C) = (X)
	(&) COMPARE	(C) = (Y)
	(&) VALUE	(C) > 0.30

ALORS SAME-PROD-METHOD (A) <— (B)

In SNARK, uncertainty of data and knowledge is handled by assigning coefficients of reliability, from zero to one, to the data or rule statements. While representing and calculating uncertainty in numerical form may not be epistemologically the most adequate representation of confidence (Huggett & Baker 1985), it can be useful if taken only as an indication of uncertainty. The initial prototype of VANDAL does not employ the uncertainty coefficients.

The SNARK inference engine represents a control strategy that evokes domain-dependant procedural knowledge to be applied to the data base; in principle it can use a forward, backward or mixed chaining search strategy. In VANDAL a forward chaining strategy with backtracking was employed, in which the search moved from the antecedent of the rule to its consequent.

SNARK, written in PASCAL VS and implemented on IBM 370/168, NAS 90 and IBM NAT, allows the use of some 500 rules and 10 000 data statements. (Today, SNARK is also available in an MSDOS version.)

VANDAL's knowledge base contains 12 production rules as well as fact base editing and output producing rules. The level of the knowledge representation in the first VANDAL prototype is 'shallow': the system contains only enough knowledge to perform a particular task rather than a complete theory of the domain. The data base for the Zagros ceramic study, used to develop and test VANDAL, contains some 380 data statements. The nature of the data is 'static'; *i.e.*, all constraints of the problem are specified before the session and it is assumed that the problem does not change during the session.

The results of a VANDAL run give, for a group of ceramics, (1) a method of production (same as or different from another ceramic group), (2) a geographic origin (local or imported from one site to another, or from a third source), (3) a continuity of production method through two consecutive periods, and (4) a continuity of local production, or import or trade, through two consecutive periods. An example of a result from a VANDAL run is:

CERAMICS MADE USING SAME PRODUCTION METHOD

WARE	SITE	PERIOD	WARE	SITE	PERIOD
HARD	SEH GABI	MC1	HARD	GODIN	MC1
BLACK-ON BUFF	SEH GABI	MC1	BLACK-ON BUFF	GODIN	MC1
BLACK ON BUFF	SEH GABI	MC1	UNTEMPERED	SEH GABI	MC1

For each of the findings, the system explicitly states the information that was used to obtain those findings.

```
A PARTIR DES FAITS CONNUS SUIVANTS:
(On the basis of the following known facts:)
      NATURE          K-SG-HARD-P1       WARE
      NATURE          K-GD-HARD-P1       WARE
      MISCLASS-R-1    VALUE              0.05
```

```
AU MOYEN DE LA REGLE SAME PRODUCTION METHOD
(using the rule SAME PRODUCTION METHOD)
JE DEDUIS LE(S) FAIT(S)
(I deduce the fact(s))
SAME PRODUCTION METHOD HARD  SEH GABI  MC1  HARD  SEH GABI  MC2
```

The system's interaction with the user is in the imperative mode; the system provides information to the user that cannot be rejected, but the knowledge base can easily be verified, changed or expanded.

24.3 Conclusions

VANDAL is a primitive expert system: it employs a 'shallow' level of knowledge, represented by production rules, to perform a simple classificatory task. It centers on a narrow provenance problem and it very much simplifies certain aspects of technical information used in provenance determination. In spite of its unsophisticated nature, its construction was very time consuming. Some of these limitations are specific to VANDAL, some to SNARK, and some to expert systems in general. However, VANDAL has proven to be a very useful tool for assembling, organizing and incorporating knowledge from various disciplines (namely archaeology, materials science, instrumental chemistry and data analysis), and performing a multidisciplinary tasks. It is hoped that further improvements, in areas such as (1) development of a more appropriate interface between statistical/numerical aspects of the provenance determination and the system, (2) selection of more appropriate methods of knowledge and uncertainty representation, and (3) incorporation of other types of information (stylistic, decorative, technical, morphological), will all lead to the development of a fully-functioning expert system for archaeological-archaeometric provenance determination.

Acknowledgments

The authors wish to thank Professor Louis D. Levine (Royal Ontario Museum) for providing archaeological expertise, and Monique Renaud (LISH, Paris) for her assistance in constructing the VANDAL system. Support for this study from CNRS-ER 315, an NSERC postdoctoral fellowship to V. Vitali, and an NSERC grant to U.M. Franklin are gratefully acknowledged.

References

CERRI, S. A., P. LANDINI, & M. LEONCINI 1987. "Cooperative Agents for Knowledge-based Information systems: Dialogue about the Archaeology of Rome", *Applied Artificial Intelligence*, 1: 1–24.

DORAN, J. 1986. "Formal methods and archaeological theory: a perspective", *World Archaeology*, 18: 21–37.

DORAN, J. 1988. "Expert systems and archaeology: what lies ahead?". *in* Ruggles, C. L. N. & Rahtz, S. P. Q., (eds.), *Computer and Quantitative Methods in Archaeology 1987*, International Series 393. British Archaeological Reports, Oxford.

GARDIN, J.-C. 1980. *Archaeological constructs.* Cambridge University Press, Cambridge.

GARDIN, J.-C. *et al.* 1987. *Systèmes experts et sciences humaines: Le cas de l'archéologie.* Eyrolles, Paris.

HUGGETT, J. & K. G. BAKER 1985. "The computerised archaeologist: the development of expert systems", *Science and Archaeology*, 27: 3–7.

LAGRANGE, M.-S. & M. RENAUD 1985. "Intelligent Knowledge-based Systems in Archaeology: A Computerized Simulation of Reasoning by Means of An Expert System", *Computers and the Humanities*, 19: 37–52.

LAGRANGE, M.-S. & M. RENAUD 1987. "Ce System SNARK". *in* Gardin, J.-C. *et al.*, (eds.), *Systemes experts et science humaines*, pp. 42–55. Hachette, Paris.

LAURIERE, J.-L. 1986. "Un langage declaratif: SNARK", *Technique et Science Informatiques*, 3: 141–172.

REICHGELT, H. & F. VAN HARMELEN 1986. "Criteria for Choosing Representation Languages and Control Regimes for Expert Systems", *The Knowledge Engineering Review*, 1: 2–17.

VITALI, V., J. W. SIMMONS, E. F. HENRICKSON, L. D. LEVINE, & R. G. V. HANCOCK 1987. "A Hierarchical Taxonomic Procedure for Provenance Determination: A Case Study of Chalcolithic Ceramics from the Central Zagros", *Journal of Archaeological Science*, 14: 423–435.