

Apprentissage automatique et Typologie.

PLATA: une expérience d'acquisition des connaissances dans le domaine de la céramique archéologique.

Marie-Salomé Lagrange

UPR 315 - CNRS
27 rue Damesme
75013 Paris

Monique Renaud

LISH - CNRS
54 Boulevard Raspail
75006 Paris

Engelbert Mephu Nguifo

CRIL - IUT Lens - Univ. d'Artois
Rue de l'université SP 16
62307 Lens cedex
E.mail: mephu@lens.lifl.fr

Résumé

Ce papier montre une aide de l'apprentissage automatique à l'élaboration d'une typologie archéologique, au stade où les classes que l'on cherche à établir sont encore mal connues. Le programme utilisé, LEGAL, est un système d'acquisition des connaissances qui permet de définir des classes sur la base de leurs exemples et de contre-exemples. Les résultats, relatifs à un corpus de céramiques du Soudan (site de Kerma), montrent la compatibilité de la discrimination obtenue avec les hypothèses archéologiques.

Mots-clés: *apprentissage, acquisition des connaissances, typologie archéologique, céramique, Soudan.*

Abstract

The authors here show that machine learning techniques can be used for designing an archaeological typology, at an early stage when the classes are not yet well defined. The program, LEGAL, is a knowledge acquisition system which uses a set of examples and counter-examples in order to discriminate between classes. Results show a good compatibility between the characterization of classes defined by the system and the archaeological hypotheses.

Key-words: *machine learning, knowledge acquisition, archaeological typology, ceramics, Sudan.*

1- Introduction

Il est admis depuis longtemps (Brachmann 1979) que les techniques d'apprentissage automatique doivent faciliter l'acquisition de connaissances; On sait aussi qu'il y a bien des problèmes pour faire comprendre et admettre par un expert humain une connaissance apprise par une machine, surtout pour la lui faire réutiliser et réviser. Cet article rend compte d'une expérimentation en apprentissage pour produire une aide à l'élaboration d'une typologie archéologique, tout en mettant en évidence de la nature des connaissances acquises. Nous cherchons dans cette étude à préciser la nature des connaissances acquises qui sont associées à l'usage de mécanismes d'apprentissage.

Les objets archéologiques traités sont ici des céramiques issues de la nécropole de la ville de Kerma (Soudan). Elles sont datées de la période Kerma Ancien et Moyen c'est-à-dire environ 2500-1800 av. J.-C. (Bonnet *et al.* eds 1990, Bonnet 1991, Privati 1982, 1986, 1988, 1990). Le problème archéologue posé, extrêmement courant, était d'affiner la classification, déjà connue, de ces céramiques, au fur et à mesure de découvertes nouvelles, et sans encore bien appréhender les "prototypes" des sous-classes que l'on recherchait. Selon l'hypothèse d'une utilisation ancienne de la zone Nord de la nécropole, et d'une utilisation progressive dans le temps en direction du Sud, ces céramiques sont datées relativement par secteur de fouille (stratigraphie horizontale).

Plusieurs systèmes d'apprentissage ont été développés ces dernières années parmi lesquels le système LEGAL (Mephu Nguifo, 1994a) que nous avons utilisé pour notre expérimentation. LEGAL s'appuie sur la structure de treillis de Galois pour générer un ensemble de caractéristiques d'un concept matérialisé par des exemples et des contre-exemples. Notre choix s'est porté sur ce système en raison de sa robustesse, de son adaptabilité au problème posé, et aux résultats obtenus par LEGAL dans d'autres applications dans le domaine de la biologie (Mephu Nguifo, 1994b).

Nous allons commencer par décrire le programme LEGAL dans la section 1. La section 2 présente les résultats expérimentaux obtenus par l'application de LEGAL au problème archéologique. Une discussion sur ces résultats fait l'objet de la section 4.

2- Le programme d'apprentissage: LEGAL

Le programme LEGAL utilisé (Mephu Nguifo, 1994a) fait partie de la famille des "agents rationnels", c'est-à-dire des systèmes automatisés d'acquisition des connaissances qui disposent de procédures pour produire et en même temps contrôler leurs propres connaissances, par le moyen d'une interaction avec un autre agent rationnel, l'utilisateur, qui joue un rôle de maître (Liquière *et al.* 1990).

LEGAL s'applique à un tableau de correspondance binaire où en lignes sont matérialisés les objets, et en colonnes les propriétés (attributs) décrivant les objets. Le système ne tient compte que des propriétés présentes. Il fonctionne en deux étapes: une étape d'**apprentissage** à partir des données qui lui sont soumises, puis une étape d'**application** ou de **décision** de la connaissance apprise (à de nouveaux objets ou bien aux mêmes objets).

2.1- Apprentissage

L'on soumet au système des **conjectures**, c'est-à-dire des exemples et (éventuellement) des contre-exemples d'un concept (d'une classe). Exemple de conjecture: la céramique issue de la nécropole de Kerma se divise en neuf classes distinctes correspondant aux secteurs de la stratigraphie horizontale. Les exemples de la classe 1 sont les céramiques décrites par le maître comme appartenant au secteur 1, avec pour contre-exemples les objets de tous les autres secteurs. Il en est de même pour chaque classe.

Soumettre une conjecture à LEGAL revient à appliquer le programme à un fichier déjà classé, présentant d'abord la description des objets de la classe que l'on cherche à discriminer des autres (ce sont les exemples), suivie de la description des objets de toutes les autres classes (ce sont les contre-exemples).

L'apprentissage est basé sur l'analyse des **ressemblances**. Le système utilise des méthodes rapides d'extraction de ces ressemblances (exploration de graphes représentant les objets décrits). Leur sélection est contrôlée par un certain nombre de critères, dont certains sont optionnels. La ressemblance devait être établie sur la base d'attributs présents à la fois dans au moins *e* exemples de la classe considérée et dans moins de *c* contre-exemples (*e* et *c* étant fixés par l'utilisateur). Les ressemblances ainsi extraites sont appelées **régularités**.

Le résultat le plus important de la phase d'apprentissage est la liste des **régularités**, suivies du nombre d'objets qui les vérifient. Les régularités sont des attributs isolés ou bien des ensembles d'attributs liés par la relation ET.

Exemple de régularité (fichier bols/jattes, céramiques de Kerma, discrimination du secteur 1, qui contient 33 objets), dont on a exigé qu'elle s'applique à 16 exemples au minimum et 17 contre-exemples au maximum):

R#1 attr: PROFSIMP PRHB1-P BEXTNONST LINCL3-N > nbre objets : 20

Ce qui signifie : "L'ensemble d'attributs: *profil simple ET Haut de panse <= Bas de panse ET extérieur de la base non stable (ovale, pointue ou ronde) ET inclinaison de la lèvre inférieure à 20°* est vérifié sur 20 objets dont au moins 16 objets sont du secteur 1, et au maximum 17 objets sont des 8 autres secteurs".

Lorsque les régularités d'un secteur comportent toutes (ou presque toutes) un ensemble d'attributs communs, nous dirons que ces attributs définissent le **prototype** du secteur. Le prototype est donc un objet fictif, qui possède certains attributs en commun avec les objets du secteur. Les objets qui possèdent tous les attributs du prototype sont des *exemples prototypiques* du secteur.

2.2- Décision

Les décisions consistent généralement à classer de nouveaux objets. La stratégie de décision est liée aux résultats de l'apprentissage, qu'elle ne peut remettre en cause. Le système classe ces nouveaux objets en testant leur comportement par rapport aux régularités. Si ces objets satisfont certains seuils - relatifs au pourcentage de régularités satisfaites, au pourcentage d'attributs utilisés dans les régularités qu'ils vérifient (par rapport à leur nombre total d'attributs), etc.. - ils sont reconnus comme **exemples** de la classe considérée. Sinon, ils sont refusés.

3- Apprentissage et décision appliqués à la céramique de Kerma

3.1- Traitement préalable des données

3.1.1- Choix du corpus

Ces données ont été étudiées, sélectionnées pour nous et entièrement décrites par Béatrice Privati, Service Cantonal d'Archéologie de Genève, Mission Archéologique de l'Université de Genève au Soudan.

Le tableau 1 montre la répartition des céramiques dans les 9 secteurs de la nécropole. L'ordre croissant des numéros de secteurs reflète l'ordre chronologique, suivant la stratigraphie horizontale. On observe que les quantités de céramiques par secteur sont peu variables, le secteur 4 étant seul assez pauvre, et le secteur 14 plus abondant.

N° secteurs	BOLS	JATTES	JARRES	POTS	TOUT
1	14	19	0	5	38
2	4	35	0	0	40
3	12	22	0	3	37
4	7	2	0	1	10
5	18	9	1	10	38
7	27	1	5	3	36
8	12	4	4	6	26
9	9	5	3	6	23
14	18	26	8	34	86
TOUT	121	123	21	68	333

Tableau 1: Répartition de la céramique dans les secteurs

(note: les *écuelles* et *bouteilles*, en très petit nombre, ont été retirés des fichiers, c'est pourquoi les totaux horizontaux ne sont pas absolument exacts)

3.1.2- Choix des attributs

Les publications décrivant des typologies archéologiques déjà établies ne sont pas démunies, en général, de commentaires abondants: ceux-ci constituent une manière de rationalisation de celles-là. Les voies qui permettent la production d'une typologie efficace (à la fois stable, résistante aux nouvelles trouvailles, et génératrice de découvertes) sont pourtant

dans l'état actuel de nos connaissances, insondables. En particulier, si l'on décide de construire une typologie en ordonnant un corpus d'objets de référence (par quelque méthode que ce soit), il n'existe pas de règles au sens strict pour la **description** initiale de ces objets.

Les attributs décrivant la céramique de Kerma ont donc été choisis de façon empirique, puisqu'il est hors de question de tout décrire, l'archéologue a tenté de traduire de façon analytique et systématique les traits qui lui paraissaient les plus significatifs au regard des objectifs, en mettant à profit toute sa familiarité avec les objets, en mobilisant le maximum de connaissances externes (d'ordre stratigraphique, historique, géographique, anthropologique).

Il est toutefois intéressant de noter que B. Privati s'est inspirée, pour la description morphologique de la céramique Kerma, du "*Code d'analyse des formes de poteries*" (Gardin *et al.* 1976). Cet ouvrage, qui fait partie d'une collection d'autres "codes d'analyse" (ou langages documentaires)¹, permet l'aide à la description de toute espèce de céramique. Les principes d'analyse y sont fort stricts: découpage invariant des objets en parties constituantes comparables, délimitation précise du sens de tous les termes descriptifs, utilisation de mesures pour exprimer les particularités morphologiques des objets, etc. Les termes proposés pour la description sont en général abrégés et codés au sens strict: *exe* pour convexe, *cave* pour concave, etc. En ce qui concerne l'élimination (raisonnée) ou le grossissement (par regroupement) de certains attributs descriptifs (à un stade de cette application), soulignons les points suivants:

- Il a fallu renoncer à garder ce que l'on peut appeler des "marqueurs archéologiques", c'est-à-dire des traits très rares mais dont on sait qu'ils indiquent certains tournants importants dans la culture matérielle (par exemple, pour la céramique Kerma Ancien, l'apparition du décor à *pois en relief*, (Bonnet *et al.* 1990, p.179, notice de B.Privati).

- Les traits redondants (corrélés statistiquement) ont été regroupés.

- Certains sous-ensembles de caractéristiques fines, apparaissant chacune trop rarement, ont été regroupés en classes plus larges. Par exemple, pour la forme extérieure de la base, les propriétés: *ovale, pointue, ronde* ont été remplacées par l'attribut "forme extérieure de la base non stable", et les propriétés: *concave, convexe, plate* par "forme extérieure de la base stable".

- Enfin, le système exigeant un codage binaire (0/1), les **dimensions** absolues des céramiques ont dû être regroupées en deux ou trois classes exprimant les notions de *petit, moyen, grand*, les seuils fixés entre ces classes étant différents pour les bols, les jattes, les jarres et les pots.

Formes et dimensions	Fabrication	Céramiques décorées
BOLJAT	FAB	DEC
JARPOT		(céramiques sans décor)

Tableau 2: Les quatre fichiers d'entrée de l'application PLATA

3.1.3- Découpage du fichier d'entrée:

Le fichier d'entrée a été scindé en quatre fichiers (tableau 2), qui ont donné lieu chacun à une application séparée de LEGAL: BOLJAT (bols/jattes), JARPOT (jarres/pots), FAB (fabrication, c'est-à-dire tournage, texture de pâte, couleur, application - il concerne la totalité des objets), et DEC (décor - il concerne la totalité des objets avec décor).

¹ Ces ouvrages ont été élaborés de 1955 à 1969 dans le cadre ou avec le concours du Centre d'Analyse Documentaire pour l'Archéologie (CNRS), dirigé par Jean-Claude Gardin. Ils ont été publiés plus tard aux Editions du CNRS.

Ce découpage, déjà adopté après des essais et erreurs au cours d'une application précédente², a été conservé moins pour des raisons pratiques d'encombrement que pour des raisons logiques:

a. Il fallait en premier lieu préserver la **comparabilité** des informations. Une partie importante de l'analyse morphologique des céramiques repose sur la forme des objets et sur les *rappports dimensionnels* entre leurs parties constituantes. Or, de ce point de vue, il n'était pas possible de comparer les bols/jattes (vases *ouverts*) et les jarres/pots (vases *fermés*).

b. Le fichier décrivant le traitement des **techniques de fabrication**, qui concerne l'ensemble de la céramique (tournage, texture de pâte, couleur, application...), était susceptible de ne pas donner de résultats significatifs, étant donné la monotonie de ces techniques à Kerma. Nous avons donc choisi de le traiter à part.

c. En ce qui concerne le **décor**, comme on le voit sur le Tableau 3 suivant, un peu plus de la moitié seulement des objets portent un décor. Si l'on avait incorporé respectivement aux fichiers BOLJAT et JARPOT les indications relatives au décor, les effectifs d'objets décorés auraient été insuffisants. Il était préférable aussi de ne pas encombrer ces fichiers d'indications relatives à l'absence de décor, lesquelles pouvaient entraîner une disparité, peut-être sans signification archéologique, entre les objets.

N° secteurs	BOL	JARRE	JATTE	POT	TOUT
1	10	0	9	5	24
2	3	0	32	0	35
3	11	0	17	1	29
4	3	0	2	0	5
5	13	1	4	4	22
7	11	2	1	2	16
8	1	1	3	6	11
9	2	1	3	6	12
14	5	0	1	11	17
TOUT	59	5	72	35	172

Tableau 3: Répartition des céramiques décorées dans les secteurs

3.2- Modalités du dialogue avec LEGAL

3.2.1- Stratégie générale

La stratégie que nous avons adoptée est la suivante:

- les résultats de la phase d'**apprentissage** ont été analysés pour établir les "prototypes" (ou, si l'on veut, le profil descriptif) des 9 secteurs;
- les résultats de la phase de **décision**, par application de LEGAL aux **mêmes objets**, ont été utilisés pour **compléter**³ les résultats de la phase d'apprentissage, en particulier:

(a) pour vérifier qu'il existait bien une discrimination entre les secteurs;

(b) pour examiner si l'on pouvait les agencer en une sériation conforme aux hypothèses archéologiques. Sur ce dernier point, nous avons supposé qu'un prototype construit par l'apprentissage du secteur S1, et qui a plus d'exemples dans le secteur S2 qu'il n'en a dans le secteur S3, est plus proche dans le temps de S2 que de S3.

3.2.2- Réglage du nombre d'exemples et de contre-exemples

Comme on l'a dit plus haut, l'utilisateur de LEGAL doit prendre certaines initiatives, en particulier, fixer lui-même le nombre d'exemples et de contre-exemples sur lequel sera opérée la discrimination entre classes au moment de l'apprentissage. Les résultats de l'application sont

² Une première approche de cette étude de cas, par des techniques d'apprentissage automatique appelées TRINITA, conçues au CRIM, CNRS, Montpellier (voir Sallantin *et al.* 1991 et Lagrange 1992) a été préalablement réalisée en 1991.

³ Nous pouvons dire que les résultats en décision ont servi à valider (par confrontation) ceux de la phase d'apprentissage. En fait, nous avons utilisé la phase de décision pour affiner nos conclusions, en particulier en ce qui concerne la sériation.

liés non seulement aux choix que nous avons faits, relatifs à la sélection du corpus, à la description des objets, au partage des données, mais aussi au réglage du nombre d'exemples et de contre-exemples. Nous indiquerons donc ici une manière de protocole par essais et erreurs, qui a été le nôtre au cours de l'application PLATA:

a. Au stade de l'apprentissage d'une classe *c* (c'est-à-dire la céramique d'un secteur donné), nous avons exigé *a priori* d'obtenir des régularités attestées dans environ 50% d'exemples de *c* (objets appartenant à *c*) et dans 10% au maximum de contre-exemples (objets appartenant aux 8 autres classes).

b. Un de nos soucis a été d'obtenir à chaque fois un nombre **utile** de régularités, c'est-à-dire: assez de régularités pour décrire un prototype de céramiques, et en même temps un nombre de régularités manipulables (pas au-delà d'une vingtaine). Cet équilibre est parfois délicat à produire (Njiwoua, 1996).

c. Cela étant, les régularités obtenues devaient aussi permettre, lors de l'étape de décision, de reconnaître un pourcentage acceptable d'exemples de la classe *c*, ainsi qu'un nombre non exorbitant de contre-exemples dans les autres classes. Les choix liés aux considérations précédentes **a** et **b** ont parfois été remis en question.

Nous donnerons maintenant un exemple précis de ce protocole dans un cas, il est vrai, un peu délicat, et qui concerne le fichier BOLJAT3 (discrimination du secteur 3, lequel comporte 34 objets). Pour illustrer le point **b**, prenons un nombre maximal de **contre-exemples** exigé constant (10), et procédons à une **augmentation** progressive du nombre exigé d'exemples, en partant de 9 exemples (tableau 4).

Nb. minimal d'exps exigé	Nb. maximal de contre-exps exigé	Nb. de régularités obtenues	Appréciation
9	10	101	Nb. de régularités trop grand
10	10	24	N. de régularités correct
11	10	4	N. de régularités insuffisant
12	10	0	Pas de résultats

Tableau 4: Nombre de régularités pour BOLJAT, secteur 3

L'on peut procéder ensuite à une **diminution** progressive du nombre maximal de contre-exemples, par exemple en partant de 12 contre-exemples, le nombre minimal d'exemples exigé restant constant (10).

En ce qui concerne le point **c**, nous avons fait varier ensuite le nombre d'exemples et de contre-exemples et procédé à des décisions (tableau 5). Nous avons finalement retenu le réglage 10/10, lequel offre le meilleur équilibre entre exemples et contre-exemples.

Réglage	Nbre de régularités	% d'exemples reconnus	% de contre-exemples reconnus comme exples
10/10	24	71	17
13/25	21	76	32
14/26	50	50	13

Tableau 5: Réglage des paramètres en fonction des résultats de la décision (BOLJAT3)

3.3- Description des résultats

3.3.1- Un exemple: BOLJAT9, traitement du fichier BOLJAT pour le secteur 9

3.3.1.1- Apprentissage

Ce fichier contient 14 exemples. Avec un réglage à 9 exemples et 20 contre-exemples, on obtient 7 régularités, comme le montre le tableau 6. Les attributs mentionnés sont les suivants:

3.3.2- Essai de synthèse des résultats d'ensemble de l'application PLATA

3.3.2.1- Bols et jattes:

Pour les secteurs 1, 2, 5, 7, 8, 9, les prototypes sont, dans l'ensemble bien marqués et les contre-exemples reflètent une continuité entre les secteurs qui est tout à fait compatible avec les hypothèses chronologiques. Le secteur 3 n'est probablement pas caractérisé par un prototype unique, le secteur 4 (pauvre en céramiques et présentant des affinités presque uniformes avec tous les secteurs) n'a pas une place bien précise dans la séquence, le secteur 14 est attiré presque également par les autres secteurs et les attire de la même façon.

Secteurs d'origine des régularités ↓	Sect.1	Sect.2	Sect.3	Sect.4	Sect.5	Sect.7	Sect.8	Sect.9	Sect.14
1	79	26	24	11	7	14	0	14	2
2	27	85	35	11	15	0	0	0	18
3	24	41	71	0	7	11	7	7	11
4	18	20	15	91	22	21	14	0	20
5	6	5	29	44	81	43	43	21	27
7	6	2,5	6	0	11	82	7	43	9
8	0	5	3	0	11	25	81	50	11
9	0	5	6	0	11	43	44	93	13
14	6	30	21	22	19	25	62	21	80

Tableau 8: Affinités entre secteurs du fichier BOLJAT

En lignes figurent les secteurs à partir desquels LEGAL a construit les régularités. En colonnes, figurent ceux où apparaissent des contre-exemples reconnus comme exemples. Aux intersections apparaissent les pourcentages d'objets (contre-exemples) reconnus comme exemples sur la base de ces régularités. Ainsi, par exemple, à l'intersection de la ligne "secteur 1" et de la colonne "Sect.7", le chiffre 14 signifie que 14% des céramiques du secteur 7 ont été considérées par LEGAL comme des exemples du secteur 1.

L'on peut ainsi évaluer d'une part la capacité des "prototypes" à reconnaître les objets de leur propre secteur (diagonale), d'autre part les affinités entre tous les secteurs, d'après les pourcentages par secteurs des contre-exemples reconnus comme exemples.

En résumé, le traitement de BOLJAT livre bien l'image d'une sériation non contradictoire avec l'hypothèse chronologique. Enfin, du point de vue de l'évolution des attributs au cours du temps, elle peut être résumée ainsi:

Type de céramique: Les jattes dominent jusqu'au secteur 3, les bols ensuite.

Dimensions: La largeur maximale, la hauteur totale, le diamètre de l'embouchure ont une tendance globale à la diminution, du secteur 1 au secteur 14.

Panse: La hauteur du bas de panse semble à peu près stable, la hauteur du haut de panse également. Une inclinaison du haut *convergente* domine les derniers secteurs, la forme du haut changeant de *concave* à *convexe*. Les rapports haut/bas de panse (dans l'ensemble $H \leq B$) sont à peu près stables. Le profil est simple d'abord, puis sinueux, enfin simple à nouveau.

Base (forme de l'extérieur): Elle évolue de "non stable" (ovale, pointue ou ronde) à "stable" (concave, convexe ou plate).

Secteurs d'origine des régularités ↓	Sect.1	Sect.5	Sect.7	Sect.8	Sect.9	Sect.14
1	100	9	25	30	11	5
5	20	91	0	30	33	0
7	0	0	100	0	11	9
8	40	9	25	90	22	2
9	20	36	37	20	89	2
14	0	18	12	20	11	90
Nombre d'objets par secteur ⇒	5	1	8	10	9	42

Tableau 9: Affinités entre secteurs du fichier JARPOT

3.3.2.2- Jarres et Pots:

La discrimination entre les profils céramiques des secteurs, obtenue à partir du fichier JARPOT, est bonne du point de vue de l'apprentissage automatique (tableau 9). L'essai de restitution de la séquence chronologique est moins convaincant. Une tentative d'explication de cet échec repose sur la faiblesse des effectifs des secteurs 1 à 9.

En ce qui concerne l'évolution des attributs au cours du temps, Il n'est pas certain que les "attributs forts" sur lesquels repose cette discrimination aient un intérêt archéologique.

3.3.2.3- Fabrication:

Les techniques de fabrication ne permettent pas d'effectuer une discrimination entre les secteurs.

3.3.2.4- Décor:

D'une manière générale, les profils descriptifs des différents secteurs sont difficile à distinguer. On note que les motifs jouent un rôle effacé dans la caractérisation des céramiques, au profit des techniques de décor. Dans beaucoup de cas (secteurs 3 à 9), les attributs dits "forts" sont les mêmes d'un secteur à l'autre. La discrimination entre secteurs est à peu près satisfaisante, sauf pour le secteur 1, qui se laisse attirer par presque tous les autres.

L'essai de restitution de la séquence chronologique est cependant bon. On remarque que le secteur 14, qui avait tendance à perturber la sériation dans le fichier BOLJAT, a bien sa place ici en fin de séquence (tableau 10).

Secteurs d'origine des régularités ↓	Sect.1	Sect.2	Sect.3	Sect.4	Sect.5	Sect.7	Sect.8	Sect.9	Sect.14
1	67	28	30	0	0	6	9	0	0
2	20	75	40	17	9	6	0	0	0
3	36	30	67	17	50	31	9	0	0
4	0	8	13	83	23	6	9	8	0
5	0	8	10	17	85	31	18	0	0
7	16	3	17	17	18	75	9	25	6
8	16	5	10	33	14	6	73	42	6
9	28	5	10	0	27	37	54	67	6
14	8	0	0	0	0	0	27	16	82

Tableau 10: Affinités entre secteurs du fichier DECOR

4. Discussion

Il existe bien des manières d'évaluer des résultats tels que les nôtres. Bien entendu, c'est à l'archéologue de décider si cette évolution des profils descriptifs des céramiques est pertinente ou non, c'est-à-dire si la caractérisation des prototypes n'est pas trop pauvre, si les attributs qui ressortent ont un sens du point de vue archéologique. Une validation pratique - par le classement, automatique, mental, ou manuel, d'autres céramiques de la même famille à l'aide de l'"expertise" construite par LEGAL - pourrait aider à la décision.

Cette décision n'est, en effet, jamais aisée. En effet, si les résultats sont, localement ou globalement, sans signification sur le plan archéologique, où placer l'erreur (ou les erreurs) de montage? Le traitement automatisé des données archéologiques (et historiques en général) repose sur une chaîne d'opérations assez longue, dont on a évoquée plus haut quelques aspects (voir §3.1 à 3.2), avec ses stratégies plus ou moins *ad hoc*, ses choix plus ou moins expéditifs, et où trouvent leur place bien des approximations, coupables ou non.

Dans le cas de LEGAL, dans la mesure où la liberté de l'utilisateur est grande, elle soulève des questions qui ne se posent pas à l'utilisateur de programmes plus classiques. En particulier, est-il légitime d'ajuster les seuils en fonction du comportement de chaque secteur, comme nous avons fait? Est-on en droit de comparer les résultats obtenus pour tous les

prototypes alors que certains sont construits sur un score sévère, et que d'autres donnent des résultats trop pauvres si l'on essaie ce même score?

En ce qui concerne les "affinités" entre secteurs, que nous avons repérées uniquement sur la base du nombre de contre-exemples reconnus comme exemples, c'est une simplification peut-être outrancière des informations apportées par le système. En effet, si l'on considère un couple de secteurs, LEGAL apporte en réalité deux informations non symétriques (non équivalentes) sur leur proximité/affinité (voir tableau 8). Soient par exemple les secteurs 5 et 7. Le prototype construit à partir du secteur 5 reconnaît comme exemples 43% du secteur 7, tandis que le prototype construit à partir du secteur 7 ne reconnaît comme exemples que 11% du secteur 5. Cette dissymétrie s'explique par le fait que LEGAL traite **séparément** chacun des fichiers décrivant un concept et ses contre-exemples (un secteur opposé à tous les autres). Sauf dans un cas d'école difficile à concevoir, il n'y a aucune raison de s'étonner d'une certaine dissymétrie. En effet la définition des exemples et des contre-exemples d'un concept est extérieure au système: il peut exister, dans ces conditions, non seulement des exemples du concept A qui ne vérifient pas les arguments en sa faveur, mais aussi des exemples du concept B qui vérifient les arguments en faveur de A. Cependant, n'avoir guère tenu compte de cette dissymétrie peut faire partie des biais introduits dans la fiabilité des résultats.

Pour tenter d'évaluer d'une autre façon nos résultats, nous les avons confrontés à ceux que l'on peut obtenir par des méthodes plus classiques de discrimination et de sériation. Nous avons appliqué au fichier BOLJAT les programmes suivants:

- le package de sériation de l'Université de Bonn (I. Scollar)
- le programme DELTA de sériation par diagonalisation de matrice (méthode des barycentres) de B. Deloche
- un programme d'attribution automatique d'objets inconnus à des classes connues utilisé en histoire de l'art: SYSTEX (Deloche 1992, Deloche *et al.* 1993). SYSTEX explore un corpus d'individus connus et en extrait deux types de "règles": **toutes les propriétés qui sont exclues, et toutes les relations d'implication** entre les propriétés qui sont présentes. Les objets inconnus sont attribués à une classe s'ils n'enfreignent aucune des règles de cette classe.

Les deux premiers programmes n'ont révélé aucune sériation compatible avec les hypothèses archéologiques. SYSTEX ne permet pas d'extraire les profils descriptifs des classes, mais la sériation que l'on peut induire des résultats de SYSTEX est cependant tout à fait comparable aux résultats de LEGAL.

Vis à vis des deux programmes de sériation, l'un des facteurs de la meilleure adéquation de LEGAL au problème de la recherche de classes ou types d'objets est qu'il tient compte (comme SYSTEX) des hypothèses sur les classes.

Par ailleurs, la supériorité des résultats de LEGAL sur ceux de SYSTEX tient en partie à une plus grande souplesse dans la définition des classes, laquelle autorise exemples et contre-exemples. Pour l'utilisateur, c'est aussi une façon plus "naturelle" de travailler.

5. Conclusion

Nous avons donc au travers de ce travail pu examiner les concepts de LEGAL qui ont permis une acquisition de connaissances, et les difficultés liées à son usage. Ces difficultés relèvent principalement du biais d'apprentissage. Nous avons par conséquent entrepris un travail d'aide au choix des paramètres d'apprentissage afin de pallier ces difficultés d'usage de LEGAL (Njiwoua, 1996).

Malgré tous les biais difficilement appréciables qui ont été introduits au cours de notre dialogue avec LEGAL, et bien qu'aucun système automatisé ne puisse, pour le moment,

apporter de "preuve" de la véracité des hypothèses archéologiques, nous pensons que cette application apporte quelque poids aux hypothèses actuelles (provisoires, rappelons-le) de l'archéologue, et des suggestions pour améliorer la typologie.

Remerciements: Ce travail a partiellement bénéficié du financement du contrat de plan Etat/Nord-Pas-de-Calais dans le cadre du projet Ganymède.

Bibliographie

- Bonnet, C. *et al.*, eds - 1990 - **Kerma, Royaume de Nubie**. Catalogue de l'exposition organisée au Musée d'art et d'histoire de Genève, 14 juin-25 novembre 1990. Edition C. Bonnet *et al.*, Genève.
- Bonnet, C. - 1991 - Les fouilles archéologiques de Kerma (Soudan). Rapport préliminaire sur les campagnes de 1988-1989, 1989-1990 et de 1990-1991, dans: *Genava*, n. s., XXXIX, p. 5-20.
- Brachmann, R.J. - 1979 - On the Epistemological Status of Semantic Networks, p3-50, dans N.V. Findler ed.: **Associative Networks: Representation and Use of Knowledge by Computers**, New-York, Academic Press.
- Deloche, B. -1992 - **Une esthétique expérimentale**. LAMPAS, Université de Lyon III, 14 avenue Berthelot, 69363 Lyon Cedex 07.
- Deloche, B., C. Lahanier, M. Jeanlin - 1993 - Attribution automatique des figurines au moyen d'un système expert. Pp. 281-296, dans (C; Brémont, M. Jeanlin et C. Lahanier (eds): **Les figurines en terre cuite gallo-romaines**. Documents d'archéologie française, N° 38, Editions de la Maison des Sciences de l'Homme, Paris.
- Gardin, J.-C. *et al.* - 1976 - **Code pour l'analyse des formes de poteries**. Editions du CNRS, Paris.
- Lagrange, M.-S. - 1992 - Symbolic data and numerical processing: a case-study in Art History by means of automated learning techniques. Pp. 330-356, dans Gardin, J.-C. et C. Peebles, eds, **Representations in Archaeology**. Indiana University Press, Bloomington.
- Lehmann, F.W. - 1992 - Semantic networks, p1-50, dans F.W. Lehmann ed.: **Semantics Networks in Artificial intelligence**, Pergamon Press.
- Liquière, M., Mephu Nguifo E., Py, M., Sallantin, J. - 1990 - La construction des Théories Semi Empiriques (TSE): les agents rationnels. Pp. 51-93 dans: **Actes des 2ièmes Journées d'Acquisition de Connaissances**, Lannion, Avril.
- Mephu Nguifo, E. - 1994a - Une nouvelle approche basée sur le treillis de Galois, pour l'apprentissage de concepts, dans: **Revue Mathématiques, Informatique et Sciences Humaines**, numéro 124, Février.
- Mephu Nguifo, E. - 1994b - Galois Lattice: A Framework for Concept Learning, dans: **Proceedings of 6th Intl. Conference on Tools with AI**, New Orleans (LA), November 6-9th, 461-468, IEEE Press.
- Njiwoua, P., Mephu Nguifo, E. - 1996 - Back from Experimentation: A study of learning biais in LEGAL, dans: **Technical Report 25**, Centre de Recherche en Informatique de Lens, Université d'Artois, June.
- Privati, B. -1982 - Nouveaux éléments pour une classification de la céramique du Kerma Ancien, *Genava*, n. s., XXX, p. 27-36.
- Privati, B. -1986 - Remarques sur les ateliers de potiers de Kerma et sur la céramique du Groupe C. *Genava*, n. s., XXXIV, p. 21-24.
- Privati, B. -1988 - La céramique de l'établissement pré-Kerma, *Genava*, n. s., XXXVI, p. 21-24.
- Privati, B. -1990- Les ateliers de potiers et leur production. Pp. 121-132 dans (C. Bonnet *et al.* eds): **Kerma, Royaume de Nubie**. Catalogue de l'exposition organisée au Musée d'art et d'histoire de Genève, 14 juin-25 novembre 1990. Edition C. Bonnet *et al.*, Genève.
- Sallantin, J., J.-J.-Szczechniarz, C. Barboux, M.-S. Lagrange, M. Renaud -1991- Théories semi-empiriques: conceptualisation et illustrations. **Revue d'Intelligence Artificielle**, Vol. 5, N°1, pp. 9-67.
- Wille, R. - 1992 - Concept lattices and conceptual knowledge systems, p493-516, dans F.W. Lehmann ed.: **Semantics Networks in Artificial intelligence**, Pergamon Press.