

# Le traitement des données archéologiques, lieu de rencontre de la sémiologie graphique et de l'analyse statistique

Bruno Desachy

▶ **To cite this version:**

Bruno Desachy. Le traitement des données archéologiques, lieu de rencontre de la sémiologie graphique et de l'analyse statistique. journée d'étude "Jacques Bertin et le Laboratoire de Graphique : l'héritage de la représentation des données en sciences sociales", projet de recherche Design graphique et recherches en sciences sociales; EHESS, Nov 2017, Paris EHESS, France. halshs-03962054

**HAL Id: halshs-03962054**
**https://shs.hal.science/halshs-03962054**

Submitted on 29 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Le traitement des données archéologiques, lieu de rencontre de la sémiologie graphique et de l'analyse statistique

Bruno Desachy (UMR 7041 ArScAn)

*intervention dans la journée d'étude "Jacques Bertin et le Laboratoire de graphique : l'héritage de la représentation des données en sciences sociales" - 21 novembre 2017 – EHESS 105 bvd Raspail 75006 Paris – amphi François Furet) - http://retrospective-bertin.ehess.fr/ -*

**Résumé :**

À la différence d'autres sciences humaines, c'est surtout le traitement graphique des données, plus que les applications cartographiques, que les archéologues ont retenu de la sémiologie graphique de Jacques Bertin. Ils y étaient prédisposés par une tradition propre de recherches de sériations chronologiques au moyen de permutations matricielles. Cette importance du traitement graphique des données conforté par les apports de Bertin a fait de l'archéologie, à la fin du siècle dernier, un lieu de confrontation mais aussi de dialogue entre sémiologie graphique et méthodes statistiques multidimensionnelles alors nouvellement apparues en sciences humaines, avec des développements qui se poursuivent encore actuellement.

**Synopsis :**

La sémiologie graphique de J. Bertin est le plus souvent connue des chercheurs en sciences humaines, en particulier des géographes et des historiens, pour ses applications en cartographie et représentation spatiale. Dans le cas de l'archéologie, elle a eu une influence notable (au moins auprès des archéologues français), mais plutôt centrée sur le traitement des données à des fins principalement chronologiques.

Cela s'explique par une tradition de recherche axée sur la typologie et la sériation (c'est à dire l'ordonnancement chronologique des caractères typologiques), qui a assez tôt recouru à des traitements graphiques « pré-bertiniens » de tableaux en présence/absence croisant ensembles d'objets (au sens large : objets, structures, sites...) et caractères descriptifs de ces objets. La diagonalisation graphique de ces matrices est recherchée par permutation manuelle des lignes et des colonnes, afin de visualiser les regroupements typologiques et d'interpréter leur ordonnancement comme une évolution chronologique (suivant un postulat d'évolutionnisme culturel hérité des sciences naturelles, confronté aux indications de datation par ailleurs disponibles). Cette utilisation des permutations matricielles, initiée dès la fin du 19$^{e}$ siècle en égyptologie (Petrie 1899), n'est jamais devenue un phénomène de masse, mais elle s'est maintenue, appliquée notamment depuis le milieu du siècle dernier à l'étude des cimetières de l'Âge du Fer (Hachmann 1950; Ruby 1993; Demoule 1999...). Elle s'est étendue au traitement des tableaux de comptages, en particulier sous forme de matrices graphiques basées sur les pourcentages de lignes (sériographes), mises en œuvre au début des années 1960 aux Etats-Unis pour l'étude de la céramique amérindienne (Ford 1962).

C'est dans ce contexte qu'en France, les archéologues prennent connaissance des travaux de Bertin, à partir de la parution de « Sémiologie graphique » (1967) où parmi des exemples de matrices ordonnables de nature historique, sociologique, ethnographique et économique figure un cas archéologique de typologie de chapiteaux), et plus encore de la publication de « La

Graphique et le traitement graphique de l'information » (1977) où J. Bertin développe et approfondit la présentation des procédures d'analyse graphique matricielle. L'apport de Bertin est essentiel : il est celui d'une systématisation et d'une axiomatisation des pratiques épisodiques précédentes des archéologues. En tant que tel il fut bien accueilli par une nouvelle génération d'archéologues (Olivier Buschenchutz, Jean-Paul Demoule, Patrick Périn, Alain Schnapp...), soucieux de formalisme processuel (suivant la double inspiration du structuralisme et du mouvement anglo-saxon de la « new archaeology »), mais de formation majoritairement littéraire et préférant le traitement graphique plus immédiatement perceptible aux analyses statistiques jugées trop abstraites et moins accessibles. Un dialogue fructueux se crée ainsi, plusieurs de ces jeunes archéologues venant consulter J. Bertin ou suivre ses enseignements. La thèse de P. Périn sur la chronologie des sépultures mérovingiennes (Périn 1980) en est un exemple majeur.

Le traitement graphique matriciel apparaissant comme une alternative aux outils mathématiques et statistiques, un autre dialogue s'est établi dans notre champ disciplinaire entre « graphiciens » empiriques et « matheux » calculateurs, là encore relativement tôt. Dès les années 1950, le traitement graphique manuel des données archéologiques fait l'objet aux Etats-Unis d'essais de formalisation, au moyen d'algorithmes d'ordonnancement automatisant les permutations (Robinson 1951; Meighan 1959) ; essais prolongés en France dans les années 1960 et 1970, en relation avec le laboratoire de Graphique (Elisseeff 1968; Legoux 1980). A partir de la fin des années 1960, le traitement graphique matriciel, assisté ou non, coexiste avec l'arrivée des techniques d'analyse statistique multidimensionnelle, à commencer par les classifications ascendantes hiérarchiques (convenant bien aux travaux de typologie), puis les analyses factorielles, en particulier l'analyse factorielle des correspondances développée sous la houlette de J. P Benzécri (1973), bien adaptée aux données qualitatives produites par les sciences humaines, et appliquée à l'archéologie sous l'impulsion notamment de F. Djindjian (1977). On peut dire que l'introduction de ces méthodes d'analyse des données a été, pour les archéologues pratiquant le traitement graphique, un choc culturel : à l'approche graphique, directe et empirique des phénomènes d'associations et d'oppositions entre individus et variables qui est celle de Bertin, s'opposait l'approche formelle des statistiques multidimensionnelles et en particulier des analyses factorielles, basées sur le calcul de distances et d'angles entre individus et/ou variables régissant leur projection dans un espace géométrique (le caractère multidimensionnel de cet espace n'autorisant pas une vision directe, mais une exploration par axes et plans successifs). Pourtant, des complémentarités se dégagent, faisant de l'archéologie un lieu de rencontre entre ces deux approches que sont la vision graphique de Bertin et la vision géométrique des statisticiens : à l'instar des premiers algorithmes évoqués ci-dessus, les outils d'analyse multidimensionnelle ont été employés pour automatiser l'ordonnancement des matrices graphiques (au moyen du classement fourni par une classification automatique, ou de l'ordre des coordonnées sur le premier axe d'une analyse factorielle) (Leredde et Djindjian 1980; Lesage 1990; Djindjian 1991). Le traitement graphique utilisant dans ce cas les rangs issus des distances entre individus et/ou variables et non ces distances elles-mêmes, on peut parler des matrices graphiques ordonnées comme d'une forme particulière, visuelle, de technique non paramétrique d'analyse des données. J. Bertin lui même a commenté cette rencontre entre sémiologie graphique et outils mathématiques dans un chapitre additif à la 3$^{e}$ édition (1999) de « Sémiologie graphique », reprenant notamment l'exemple de la sériation des objets mérovingiens étudiés par P. Périn. retraité à l'aide de plusieurs algorithmes.

Depuis, la révolution micro-informatique et plus encore la révolution Internet ont généré puis rendu accessibles d'énormes quantités de données (y compris dans le domaine de l'archéologie, l'afflux de données étant là soutenu par le développement de l'archéologie préventive), et ont démocratisé les outils d'analyse statistique (via notamment le langage R gratuitement accessible). On ne peut nier cependant que les outils d'analyse statistique multidimensionnelle des données se sont assez peu répandus dans notre discipline, ce qui est dommage car leur puissance est utile face à cet afflux de données produites par l'archéologie préventive ; et quant à la sémiologie graphique appliquée au traitement des données, même si elle s'est en partie maintenue chez les archéologues, et même si elle est un précurseur de l'actuel courant des « DataViz », elle n'a pas eu, plus généralement, l'audience qu'elle méritait. Ainsi, comme le remarquait S. Bonin (1998), les logiciels de type Excel ont – à côté de remarquables qualités – imposé massivement l'usage de graphiques standard qui sont le plus souvent l'exemple de ce qu'il ne faut pas faire...

Pourtant, cette rencontre entre statistiques et graphique matricielle appliquées à l'archéologie n'est pas restée lettre morte, et connaît aujourd'hui encore des développements. Nous souhaitons en particulier évoquer le cas de l'analyse des matrices de contingence, en d'autres termes le traitement des tableaux de comptages tels ceux obtenus par tris croisés des résultats de sondages (croisant par exemple les catégories socio-professionnelles des personnes sondées avec leurs types de pratiques culturelles ou leurs préférences politiques). Dans le cas de l'archéologie, ce même type de tableau permet d'exploiter les inventaires issus de fouille, en croisant les types d'objets ou d'éléments recueillis avec les couches ou structures archéologiques de provenance, c'est à dire en se fondant sur la relation contexte / objet pour mettre en évidence des phénomènes temporels, spatiaux ou fonctionnels (à la nuance près que le décompte d'élément par type et par contexte de provenance n'est pas une opération simple – comme toute statistique appliquée à l'archéologie : même nos individus sont en ruines ! – et nécessite le recours à plusieurs indicateurs tels le nombre de restes, le nombre minimum d'individus, etc.). Ce type de tableau, correspondant à la répartition d'un nombre total d'éléments à travers deux séries de modalités qualitatives (en lignes et en colonnes), est classiquement traité par analyse factorielle des correspondances pour analyser les variations proportionnelles traduisant des associations et oppositions entre lignes et colonnes. L'importance de ces variations est donnée par un indicateur statistique bien connu – le Khi-deux – que l'on peut décrire comme une mesure d'écart aux proportions moyennes du tableau (lesquelles correspondent à la situation statistique dite d'indépendance, où il n'y a aucune association ou opposition entre lignes et colonnes), et que l'on peut tester pour évaluer le risque que ces associations et oppositions ne soient que des variations aléatoires. Le recours à ces procédures et indicateurs statistiques chiffrés n'est pas dans la démarche ni la culture de J. Bertin ; cependant – preuve que son approche empirique est solidement raisonnée – il en retrouve les fondements dans le traitement qu'il propose pour les matrices de contingence, à savoir la « matrice pondérée ». Dans « la Graphique » (1977, p.62-63), il base en effet la construction d'une « matrice pondérée » sur la comparaison entre des valeurs de références calculées à partir des totaux du tableau (qui ne sont autres que la situation d'indépendance, appelée « répartition proportionnelle » par Bertin), et les valeurs réelles (« répartition observée »). Cette comparaison visuelle est guidée par la mise en évidence, en noir, des écarts positifs aux moyennes de lignes ou de colonnes (qui ne sont pas exactement des écarts à l'indépendance car calculés seulement sur les lignes ou sur les colonnes) et s'appuie sur les surfaces attribuées aux valeurs, calculées par rapport aux totaux de lignes et de colonnes. C'est là, au fond, le principe des écarts à l'indépendance et de l'indicateur Khi-deux (caractérisé par

une pondération par les totaux de lignes et de colonnes), dont Bertin trouve ainsi une traduction graphique. La rencontre entre graphique et outils statistiques n'est ici plus seulement à la marge (des tableaux de données...) pour automatiser l'ordonnancement des lignes et des colonnes, mais se situe plus au cœur de l'analyse statistique, dans l'appréciation de l'importance des phénomènes observés.

Bien que les travaux de Bertin et de son équipe sur le traitement des données soient comme on l'a dit restés trop peu connus, ce point précis de l'acuité statistique des « matrices pondérées » n'a heureusement pas échappé à certains chercheurs : ainsi le sociologue et statisticien Ph. Cibois, l'un des promoteurs de l'analyse factorielle en sciences humaines et qui a lui même développé des outils informatiques de visualisation (Cibois 2003), ou l'archéologue et préhistorien R. Chenorkian (Chenorkian 1996). S'inscrivant dans cette lignée, l'auteur de ces lignes a travaillé sur la représentation visuelle des écarts positifs et négatifs à l'indépendance sur les matrices graphiques ordonnables et en particulier les sériographes utilisés pour le traitement de données archéologiques (Desachy 2004; 2016).

Un dernier mais pas moindre point à noter concernant cette rencontre entre archéologie, sémiologie graphique et statistiques, est l'intérêt pédagogique de la seconde : l'expérience, au moins dans le cas du traitement des données archéologiques, montre qu'elle facilite l'accès aux outils statistiques des doctorants et chercheurs « non-matheux ». Ceux-ci peuvent en effet en quelque sorte « étalonner » la vision analytique, géométrique, des phénomènes explorés axe par axe donnée par l'analyse factorielle, au moyen de la vision synthétique apportée par la graphique matricielle ; ils peuvent ainsi progressivement mieux apprivoiser cette vision géométrique au départ quelque peu abstraite. Ce rôle de passerelle entre humanités et outils mathématiques suffirait à lui seul, si il en était besoin, à prouver l'importance et l'actualité des travaux de Jacques Bertin.

**Références citées :**


Benzécri, Jean-Paul, éd. 1973. *L'analyse des données (1 La taxinomie 2 L'analyse des correspondances)*. 2 volumes : 1 La taxinomie 2 L'analyse des correspondances vol. Paris: Dunod.

Bertin, Jacques. 1967. *Sémiologie graphique les diagrammes, les réseaux, les cartes*. Réimpression 1999. les réimpressions des éditions de l'EHESS. Paris: Ed. de l'EHESS.

———. 1977. *La graphique et le traitement graphique de l'information*. Nouvelle bibliothèque scientifique. Paris: Flammarion.

Bonin, Serge. 1998. « Le développement de la graphique de 1967 à 1997 ». *Bulletin du comité français de cartographie*, nº 156: 17-21.

Chenorkian, Robert. 1996. *Pratique archéologique statistique et graphique*. Paris, Lattes: Éd. Errance et Association pour la diffusion de l'archéologie méridionale.

Cibois, Philippe. 2003. *Les écarts à l'indépendance. Techniques simples pour analyser des données d'enquêtes*. Méthodes quantitatives pour les sciences sociales, collection de livres en ligne dirigée par Alain Degenne et Michel Forsé. revue Sciences Humaines. http://www.scienceshumaines.com/textesInedits/Cibois.pdf.

Demoule, Jean-Paul. 1999. « Chronologie et société dans les nécropoles celtiques de la culture Aisne-Marne du VIe au IIIe siècle avant notre ère. » *Revue Archéologique de Picardie*, nº Numéro spécial 15: 406.

Desachy, Bruno. 2004. « Le sériographe EPPM : un outil informatisé de sériation graphique pour tableaux de comptages ». *Revue archéologique de Picardie* 2004 (3/4): 39-56. https://doi.org/10.3406/pica.2004.2396.

———. 2016. « Explographe 1.0 – application de traitement graphique de tableaux de données sous LibreOffice Calc ». *ArchéoFab - Archéologies du Bassin Parisien  fabrique numérique pour l'archéologie* (blog). 2016. https://abp.hypotheses.org/5413.

Djindjian, François. 1977. « Burin de Noailles, burin sur troncature et sur cassure : statistique descriptive appliquée à l'analyse typologique ». *Bulletin de la Société préhistorique française. Comptes rendus des séances mensuelles* 74 (5): 145-54. https://doi.org/10.3406/bspf.1977.8497.

———. 1991. *Méthodes pour l'archéologie*. U. Archéologie. Paris: A. Colin.

Elisseeff, Vadime. 1968. « De l'application des propriétés du scalogramme à l'étude des objets ». In *CNRS*. Vol. 107-120.

Ford, James A. 1962. *A quantitative method for deriving cultural chronology*. Technical manual 1. Washington: Pan American Union. https://catalog.hathitrust.org/Record/000291635.

Hachmann, Rolf. 1950. « Das Gräberfeld von Rondsen (Radz), Kreis Gaudenz (Grudziadz), und die Chronologie des Spätlatènezeit im östlichen Mitteleuropa ». *Archaeologia Geographica*, nº 1 (1951): 79-96.

Legoux, René. 1980. « Le recours à l'informatique : la chronologie relative par permutation matricielle automatique ». In *La datation des tombes mérovingiennes , historique, méthodes, applications*, par Patrick Périn, 138-55. Genève: Droz.

Leredde, Henri, et François Djindjian. 1980. « Traitement automatique des données en archéologie ». *Dossiers de l'archéologie*, nº 42 (avril): 52-69.

Lesage, Jean. 1990. *Archéologie et statistique - application des analyses factorielles*. Archives d'Ecologie Préhistorique 10. Toulouse: Ecole des Hautes Etudes en Sciences Sociales.

Meighan, Clement W. 1959. « A New Method for the Seriation of Archaeological Collections ». *American Antiquity* 25 (2): 203-11. https://doi.org/10.2307/277439.

Périn, Patrick. 1980. *la datation des tombes mérovingiennes -historique, méthodes, applications*. centre de recherches d'histoire et de philologie de la IVe section de l'Ecole Pratique des Hautes Etudes - V - Hautes Etudes Mediévales et Modernes 39. Genève: Droz.

Petrie, W. M. Flinders. 1899. « Sequences in Prehistoric Remains ». *The Journal of the Anthropological Institute of Great Britain and Ireland* 29 (3/4): 295-301. https://doi.org/10.2307/2843012.

Robinson, S.W. 1951. « A Method for Chronologically Ordering Archaeological Deposits ». *American Antiquity* 16 (4): 293-301. https://doi.org/10.2307/276978.

Ruby, Pascal. 1993. « Sériations et traitements de données archéologiques avec le logiciel Microsoft Excel sur Macintosh (ou compatible IBM) ». *Archeologia e calcolatori*, nº 4: 81-111.