

PACO, une fonction pour classer les données archéologiques sous R : quel avenir à cet outil?

Discussion de projet

résumé

Marie Philippe (ANTEA-Archéologie, membre associé à l'UMR 6298 ARTEHIS)

La fonction PACO (Partitioning Archaeological Chaînes Opératoires) a été développée dans le cadre d'un travail doctoral s'intéressant aux chaînes opératoires de fabrication des céramiques (Philippe, 2018).

L'identification des chaînes opératoires nécessite de plus en plus la prise en compte d'une grande quantité de variables qualitatives, permettant de détailler plus encore les manières de faire des potiers. Paradoxalement, les outils de traitement et de modélisation adaptés à ces données en archéologie sont encore peu développés. En effet, les bases de données comprennent de nombreuses valeurs indéterminées, dues entre autres aux conditions de conservation, tandis que la variabilité à documenter peut être très grande. Dans ce cas, identifier des chaînes opératoires à partir des multiples variables techniques enregistrées est un travail long et fastidieux s'il n'est pas automatisé.

Plusieurs problèmes méthodologiques étaient donc posés dans le cadre du travail doctoral :

- Comment identifier rapidement et efficacement les chaînes opératoires à partir d'un jeu de données ?
- Dans le cas où de très nombreuses chaînes opératoires différentes sont identifiées, faut-il regrouper les plus similaires en classes ? Comment définir cette similarité ?
- Comment représenter les résultats ?

Pour les solutionner, une fonction portant l'acronyme PACO a été construite sous R. Inspirée des recommandations du manuel de technologie céramique de Valentine Roux (2016), il s'agit d'un processus de partitionnement itératif traitant une à une les variables comportant le plus de valeurs identiques. La visualisation sous forme d'arbre est réalisée via le package `data.tree`. Cette manière de hiérarchiser les variables en fonction de la donnée disponible, et non pas de leurs enchaînements relatifs dans le processus de fabrication ni d'un a priori issu des observations ethnographiques, est une première réponse aux problèmes de gestion des indéterminations dans les assemblages archéologiques. La fonction organise les données et met en relief le degré de normalisation et de variabilité technique de l'assemblage.

Depuis 2018, plusieurs chercheurs en anthropologie des techniques ont manifesté leur intérêt pour un tel outil. En réalité, cette méthode de traitement automatisé de grands jeux de données lacunaires serait transférable dans un large panel de disciplines et thèmes de recherche liés à l'archéologie, où aucun outil similaire n'existe. Pour combler ce manque, il serait utile que PACO soit amélioré et mis à

disposition de la communauté. Cette communication permettra de présenter l'outil tel qu'il existe dans son état actuel, de discuter avec les participants des possibilités de test et de développement, lui permettant de s'adapter à des problématiques et des jeux de données plus variés, et d'inventer les conditions concrètes de mise en œuvre de ce projet.

références citées

Philippe 2018 : PHILIPPE M. (2018) – Les traditions techniques céramiques de la vallée du Rhin supérieur entre X^e et VIII^e siècles av. J.-C. : essai d'un outil automatisé de partitionnement de chaînes opératoires (PACO), thèse de doctorat, Université de Bourgogne Franche-Comté, Dijon, 515 p.

Roux 2016 : ROUX V. (2016) – Des céramiques et des hommes : décoder les assemblages archéologiques, Nanterre, Presses universitaires de Paris Ouest, 415 p.