

APPLICATION DU *MACHINE LEARNING* POUR L'ANALYSE TECHNO-STYLISTIQUE DE L'ART MOBILIER MAGDALÉNIEN.

Contexte :

Les manifestations graphiques préhistoriques sont le reflet de la culture et de l'organisation sociale des communautés, de leurs croyances religieuses et de leur rapport au monde. Ces manifestations graphiques ont des caractéristiques propres. Le contexte de cette étude doctorale est celui du Magdalénien (21 - 14 000 BP), époque d'apogée de l'art paléolithique en Europe occidentale. Le corpus, riche de plusieurs centaines d'objets, comprendra des représentations jugées parmi les plus « conventionnelles », ou « stylisées », provenant d'une quarantaine de grands sites magdaléniens du Périgord jusqu'en Asturies. Le cadre conceptuel de la mémétique et les méthodes statistiques issues du *machine learning* permettront d'étudier la diffusion de ces figures. Dans le contexte actuel de rupture technologique due à l'Intelligence Artificielle, il est proposé, pour analyser les données et quantifier les proximités, d'appliquer et de développer différents types de méthodes statistiques d'apprentissage automatique. Avec le développement de l'intelligence artificielle et du traitement du signal, le *machine learning* s'inscrit dans un contexte de renouvellement méthodologique transdisciplinaire porté par de nombreux congrès internationaux tels que AISTATS, NEURIPS, MVMML, etc. Le traitement de l'image en art paléolithique par des méthodes statistiques est un domaine novateur qui doit se développer et se structurer au sein de la communauté internationale. Ce projet doctoral est une initiative unique dans ce domaine sur le plan national. Ces développements méthodologiques seront conduits en étroite collaboration avec l'Institut de Mathématiques de Bourgogne, et établiront les bases d'un nouveau cadre méthodologique capable de retracer les influences techno-stylistique du Magdalénien, mais qui sera aussi déployable sur d'autres problématiques à l'avenir.

Objectif scientifique :

Durant le Magdalénien, les communautés humaines montrent un haut degré de structuration. Elles échangent et s'influencent sur de vastes territoires. Des réseaux de diffusion de différentes matières premières se mettent en place et irriguent des régions parfois très éloignées les unes des autres. Ces réseaux permettent de tracer les contours de territoires où des relations sociales, économiques et techniques sont partagées. Ces réseaux d'échanges et d'influence commencent à être bien documentés dans le milieu de l'artisanat des matières dures minérales et parfois animales. C'est également le cas dans le champ des comportements symboliques. Certaines figures dites "conventionnelles" - dont la composition est contrainte par des normes graphiques importantes - se retrouvent en nombre dans une multitude de sites archéologiques. Grâce au cadre conceptuel de la mémétique, c'est-à-dire du transfert de pratiques culturelles dans les sociétés humaines, il est possible de reconstituer les liens privilégiés qui reliaient les groupes magdaléniens à partir des proximités entre leurs productions.

L'étude s'appuiera sur les données directement issues de la lecture et de l'analyse typo-technologique et stylistique des pièces d'art mobilier. Le relevé des entités graphiques et le scan 3D des objets, tels que pratiqués aujourd'hui dans la communauté scientifique, constitueront un préalable méthodologique au traitement statistique de l'ensemble des observations. Les méthodes de *machine learning*, basées sur l'analyse d'image, la sélection de variables en haute dimension (LASSO, SLOPE, ...), l'algorithme des K-moyennes, les Cartes de Kohonen et le consensus de partition permettront la production d'une carte d'influence artistique. L'un des axes du projet doctoral vise à diversifier et améliorer les méthodes de *machine learning* sur des données archéologiques et expérimentales. Cette carte d'influence artistique fera par la suite l'objet d'une discussion méthodologique.

Pour ce faire, un large corpus mobilier magdalénien (supports lithiques et sur matières dures d'origine animale), composé de représentations figuratives de style conventionnel, sera mobilisé. Adossé à des inventaires actualisés récemment, ce corpus comprendra notamment des chevaux dits "macrocéphales" sur industrie osseuse en bois de cervidés (près de 300 entités graphiques), des contours découpés de têtes animales sur stylohyoïdes (environ 130 EG), des animaux schématiques vus de face sur os ou bois de cervidés (une centaine d'EG), et des représentations humaines segmentaires (plusieurs centaines d'EG). D'autres thèmes de style conventionnel pourront enrichir cette approche techno-stylistique innovante. Il est attendu la mise en évidence des mécanismes de diffusion de ces images au sein de leurs territoires culturels. De nouveaux outils pour l'étude de l'art préhistorique seront développés et participeront à la recherche dans le domaine de l'apprentissage artificiel. Les données générées par ce travail doctoral seront déposées sur des plateformes dédiées aux bases de données archéologiques ouvertes (Nakala, Huma-Num, POP, etc.) et le code des algorithmes sera déposé sur GitHub dans un souci de reproductibilité et de transparence. Les résultats de ce travail feront l'objet de communications dans des colloques en *Machine Learning* et en Archéologie, et seront publiés dans les revues internationales de rang A telles que *Paléo* et *Journal of Archaeological Sciences*.

Justification de l'approche : Par le biais d'une application inédite du machine learning à l'analyse techno-stylistique des figurations mobilières conventionnelles et de leurs supports, la thèse doit permettre d'identifier les réseaux de diffusion et les connexions existants entre les différentes régions du domaine franco-cantabrique durant le Magdalénien. La mise en œuvre de méthodes d'apprentissage artificiel permettra d'appréhender de manière moins subjective les phénomènes de convergence et de divergence d'expressions graphiques et techniques et de mieux caractériser les réseaux d'échanges parfois difficiles à cerner (paléogéographie, centralisation autour de certains points, sites d'agrégation, etc.). Nous comparerons également les réseaux de diffusion de l'art mobilier à ceux connus des matériaux lithiques et osseux.

Adéquation aux axes d'OPUS et SPHYNX : Ce projet doctoral est fondamentalement interdisciplinaire (mathématiques, informatique, archéologie, archéozoologie, art préhistorique). Il stimulera la collaboration entre plusieurs domaines de recherche rarement associés. Il permettra le développement d'algorithmes d'intelligence artificielle open-source inédites qui ouvriront à une plus grande reproductibilité des résultats dans un domaine en recherche constante de renouvellement méthodologique (OPUS : **Intelligence artificielle et réalité virtuelle**). L'attention toute particulière portée aux objets d'art mobilier méconnus, constituant le corpus, permettra de produire des données et une documentation conséquente sur ces collections patrimoniales exceptionnelles (bibliographie et histoire des collections, recherche de pièces disparues, état de conservation, etc.) permettant leur valorisation auprès de la communauté scientifique internationale et du public, notamment à travers leur numérisation en 3D (OPUS : **Numérisation et patrimoines**). Les objectifs de la thèse sont tout à la fois axés sur la production d'une nouvelle méthode d'étude du style artistique pour en explorer la variabilité formelle (SPHYNX : **Canons artistiques - normes et évolutions**) que sur la mise en évidence des structures des sociétés du passé qui sont encore détectables à travers le répertoire d'images (SPHYNX : **Constitution et préservation des patrimoines savants et Identité, communs, mémoire**).

Rôle des encadrants :

1/ **Patrick PAILLET** (MCM HDR) - UMR 7194 'HnHp', MNHN. P. Paillet est spécialiste de l'art du paléolithique supérieur en Europe et en Amérique du Sud. Son rôle d'encadrant principal sera d'accompagner, de conseiller et d'orienter le/la doctorant.e dans ses recherches au sujet des expressions graphiques préhistorique.

2/ **Patrick TARDIVEL** (MCU HDR) - UMR 5584 'IMB', Univ. de Bourgogne. P. Tardivel mène des recherches sur l'apprentissage statistique, avec des applications transdisciplinaires. Son rôle sera d'encadrer le/la doctorant.e dans ses recherches en *machine learning* et en statistiques.

3/ **Catherine SCHWAB** (Conservatrice en chef du Patrimoine) – Musée d'Archéologie National (MAN) - UMR TEMPS 8068, Univ. Paris 10 Nanterre. C. Schwab est responsable du département Préhistoire (Paléolithique et Mésolithique) du MAN. Elle fournira l'accès aux collections patrimoniales et encadrera le/la doctorant.e autour des questions relatives à l'industrie osseuse et aux considérations techniques sur la production de ces objets.

Publications majeures en lien avec le projet :

Paillet P., Paillet E., 2022. 'Le seul réel dans l'art, c'est l'art'. Réalisme, naturalisme et illusionnisme des oeuvres d'art paléolithiques. Bulletin de la Société préhistorique française, 119, 3, 421-447.

Coqueugniot H., Dutour O., Jaubert J., Paillet P. (dir.), 2023. Les sciences archéologiques à l'ère du virtuel. Actes des congrès nationaux des sociétés historiques et scientifiques, Marseille 2021, éditions du Comité des Travaux historiques et scientifiques (CTHS), 234 p.

Tardivel P., Bogdan M., 2022. On the sign recovery given by LASSO, thresholded LASSO and thresholded basis pursuit denoising, Scandinavian Journal of Statistics, 49(4), 1636-1668.

Schneider U., Tardivel P., 2022. The Geometry of Uniqueness, Sparsity and Clustering in Penalized Estimation, Journal of Machine Learning Research, 23(331), 1-36.

Schwab C., 2008. La collection Piette. Musée d'Archéologie nationale, Château de Saint-Germain-en-Laye, RMN, 127p.

Schwab C., Man-Estier E., 2016. L'Ours dans l'art préhistorique. Musée d'Archéologie nationale, Château de Saint-Germain-en-Laye, RMN, 82 p.

Profil recherché : Le/la candidat.e devra disposer d'un master en préhistoire spécialisé en art préhistorique, et d'un master en *machine learning* ou en statistiques. Il/elle devra avoir obligatoirement une expérience d'étude et de relevé d'art mobilier, rupestre et/ou pariétal. Il/elle devra par ailleurs avoir déjà mis en pratique des analyses statistiques dans ce domaine et devra connaître la panoplie actuelle des outils numériques d'enregistrement 2D et 3D (Photoshop ou équivalent, photographie, RTI, numérisation 3D). Une expérience de terrain en archéologie (participation et /ou responsabilité de fouilles) sera appréciée.