

## **Identification de l'âge et du sexe des défunt à partir du mobilier funéraire mérovingien : une approche d'apprentissage automatique**

GAËL POUX-MÉDARD

Docteur, UMR 3083 ERIC, université lumière Lyon 2

gael.poux@yahoo.fr

FRANCK ABERT

Musée Archéologique de Dijon

fabert@ville-dijon.fr

HÉLÈNE BARRAND EMAM

ANTEA-Archéologie, UMR 7044

helene.barrand@antea-archeologie.com

FANNY CHENAL

Inrap, UMR 7044

fanny.chenal@inrap.fr

THOMAS FISCHBACH

Archéologie Alsace, UMR 7044

thomas.fischbach@antea-archeologie.com

Élisa SCHNEIDER

Doctorante, UMR 7041 ArScan, Université Paris 1

Panthéon-Sorbonne

schndr.elisa@gmail.com

AMANDINE TOUZET

ANTEA-Archéologie, UMR 7044

amandine.touzet@antea-archeologie.com

### **Résumé**

L'apparition de bases de données archéologiques massives conduit à se poser la question des nouveaux outils permettant leur analyse. Les méthodes d'apprentissage automatique semblent toutes désignées pour exploiter ces nouveaux corpus. Dans cet article, nous appliquons le modèle inférentiel SIMSBM à un corpus de mobilier funéraire retrouvé dans des sépultures mérovingiennes de la région Alsace. Notre objectif consiste à dévoiler l'existence de corrélations entre ces objets et le sexe ou l'âge des défunt, évaluer leur importance, et proposer des méthodes de visualisation afin d'établir leur pertinence.

Nos résultats suggèrent l'existence de fortes corrélations entre mobilier funéraire et sexe d'un individu et la rareté, sinon l'absence, de telles corrélations avec sa classe d'âge. Ces corrélations permettent la prédiction du sexe d'un individu sur la seule base du mobilier avec près de 90% de précision, laissant entrevoir la possibilité d'élaborer à terme des outils d'annotation automatiques et fiables à destination de corpus similaires.

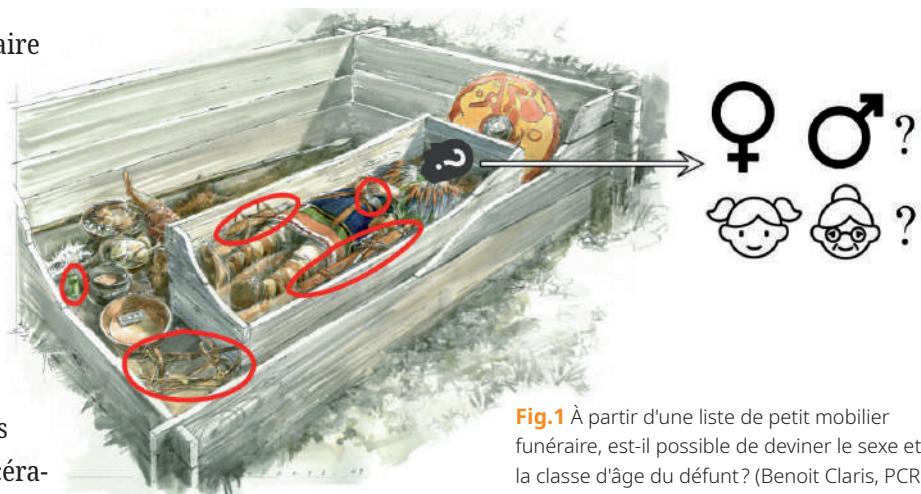
### **Mots-clés**

Mérovingien, Grand Est, mobilier funéraire, machine learning, base de données, archéologie funéraire, sexe et genre

## Introduction

Les principes de la science ouverte et la mise en commun des résultats de recherches requièrent la mise en place de nouveaux outils de collecte, de traitement et d'analyse des données. Aussi, des bases de données conséquentes commencent à émerger dans le monde des sciences humaines et, avec elles, le besoin d'outils originaux et adéquats permettant d'exploiter ces corpus inédits (**fig. 1**). Nous détaillerons ici le fonctionnement de l'un de ces outils, le SIMSBM, avant de l'appliquer à l'étude d'une problématique concrète, à l'aide d'un corpus archéologique conséquent traitant du mobilier funéraire mérovingien.

L'abondance de mobilier funéraire dans les inhumations de la période mérovingienne permet, notamment en Alsace, d'envisager une détermination du sexe basée uniquement sur des critères archéologiques, lorsque la détermination morphologique n'est pas envisageable. En effet, si certains types de mobilier sont plutôt mixtes (céramiques, couteau, peigne...), d'autres sont préférentiellement associés à des sépultures masculines (objets relatifs à l'armement : scramasaxe, *umbo*, épée, fer de lance par exemple) ou féminines (objets relatifs à la parure : collier de perles, boucles d'oreilles ou encore châtelaine). Cette approche est fréquemment utilisée en Alsace pour déterminer le sexe d'un individu en complément de la diagnose basée sur des critères exclusivement biologiques. Cette méthode n'avait jusqu'alors fait l'objet d'aucune validation systématique à grande échelle et son application s'en trouvait régulièrement discutée et remise en cause, à juste titre.



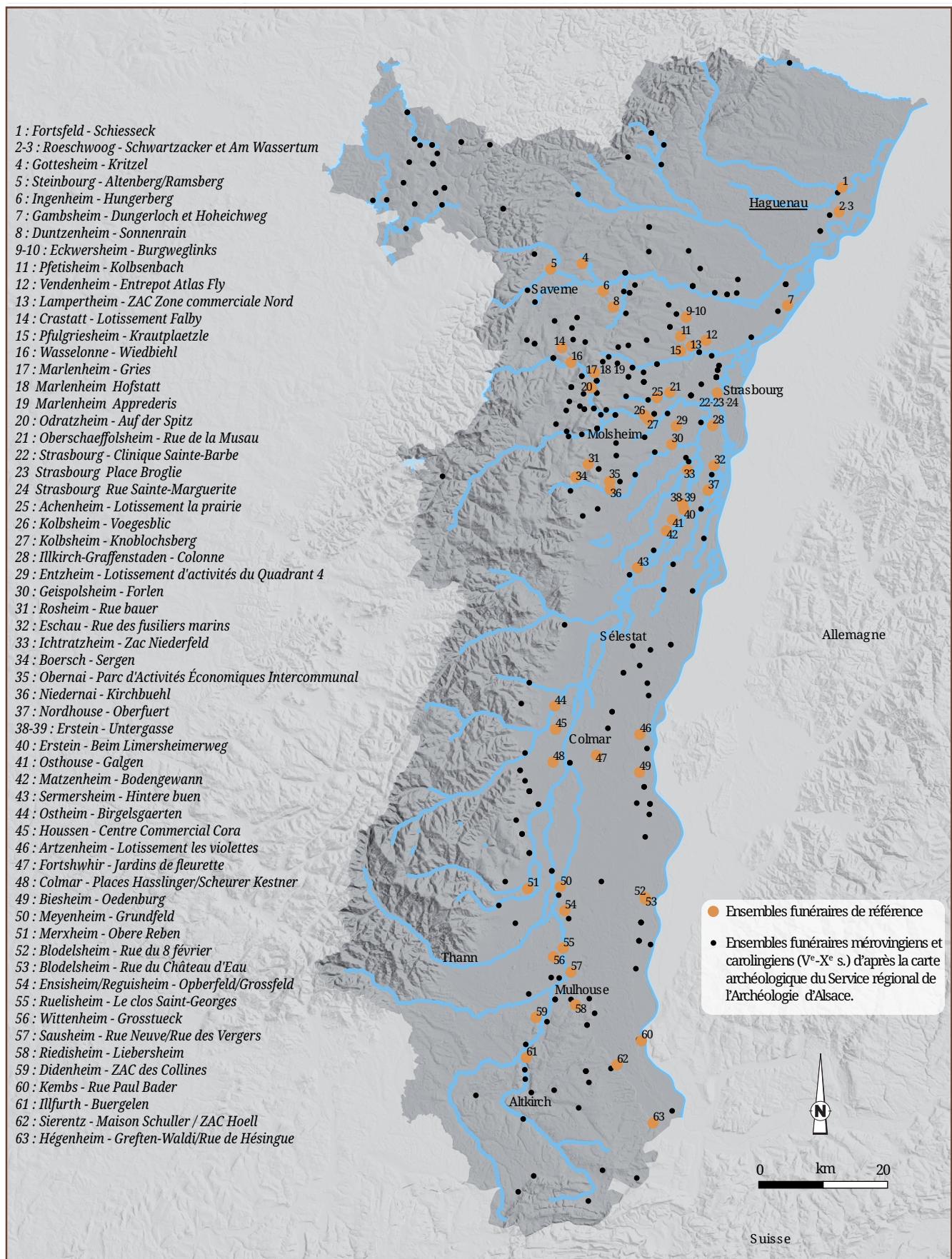
**Fig.1** À partir d'une liste de petit mobilier funéraire, est-il possible de deviner le sexe et la classe d'âge du défunt? (Benoit Claris, PCR funéraire mérovingien Alsace)

Afin d'évaluer la fiabilité de cette méthode et de tester, ou découvrir, les liens éventuels entre sexe des défunt et mobilier funéraire, un premier examen statistique des données régionales a été mis en place dans le cadre du PCR Espace et pratiques funéraires en Alsace aux périodes mérovingiennes et carolingiennes. Une première étape consistait en la création d'un corpus le plus large possible, constitué de tombes associant un dépôt funéraire d'objets jugés discriminants et des données biologiques fiables, basées sur la détermination du sexe morphologique à l'aide de la Diagnose Sexuelle Probabiliste (**MURAIL et al. 2005, BRUZEK et al. 2017**), associée pour certaines tombes à des données paléogénétiques<sup>1</sup>. L'équipe de recherche a donc créé les outils nécessaires à travers la mise en place de plusieurs bases de données, dont une était dédiée aux données biologiques concernant les individus inhumés<sup>2</sup> et une seconde, aux dépôts funéraires présents dans les sépultures<sup>3</sup>. Au cours des deux triennales du projet, ce sont un total de 2439 sépultures provenant de 58 sites qui ont été réexaminées (**fig. 2**) : les données issues des fouilles les plus anciennes ont été réévaluées et l'ensemble des données ont été compilées, permettant ainsi d'engager une analyse inférentielle à partir de données fiables et récentes (**BARRAND EMAM et al. 2021**).

1 Étude réalisée dans le cadre de l'ANR ANCESTRA dirigé par Mélanie Pruvost (UMR 5199 PACEA).

2 Coordination Fanny Chenal.

3 Coordination Franck Abert et Amandine Touzet.



**Fig. 2** Répartition des ensembles funéraires mérovingiens et carolingiens en Alsace (V<sup>e</sup>-X<sup>e</sup> s.) avec mention des sites retenus dans le corpus d'étude. (Fond de carte © IGN 2013; sites géoréférencés : extrait du SIG sous QGis. (A. Touzet/H. Barrand Emam, ANTEA Archéologie)

Au sein de la base de données dédiée aux dépôts funéraires, chacun des individus inhumés est associé à un sexe biologique (masculin ou féminin, mais parfois indéterminé), déterminé sur la base de critères morphologiques et / ou génétiques en fonction de la méthode d'identification utilisée. Il est également associé à une classe d'âge (adulte ou immature), ainsi qu'à une liste de mobilier retrouvé dans la sépulture. Du fait de la variété de ce dernier, une classification en 65 catégories a été établie (**BARRAND EMAM et al.** à paraître). Il a été choisi de classer les objets avant tout selon leur fonction : un peigne en os et un peigne en bois de cervidé appartiendront par exemple à la même classe d'objets « peigne ».

De cette base de données a été extraite une première série de tombes comprenant plus de 2000 objets répartis entre environ 250 sépultures associées à une classe d'âge, et environ 100 sépultures associées à un sexe déterminé sur la base de critères morphologiques et / ou génétiques. Ce corpus conséquent permet donc de mener une étude à l'aune de récents progrès en apprentissage automatisé. Spécifiquement, nous rechercherons l'existence de possibles corrélations entre le mobilier retrouvé dans les sépultures et le sexe et la classe d'âge des défunt.

## Motivation

Au cours du XIX<sup>e</sup> et XX<sup>e</sup> siècles, l'archéologie funéraire mérovingienne connaît un véritable essor grâce à l'explosion du nombre de découvertes de sépultures, la menant à une reconnaissance plus franche en tant que discipline indépendante (**EFFROS 2003 ; 2012**). Dans l'ensemble complexe que constitue une sépulture, le mobilier funéraire représentait alors la source d'intérêt principale des archéologues. Ce dernier servait de fondement à l'étude de l'identité des individus inhumés. De façon automatique, et ce dès les premières découvertes attestées de sépultures mérovingiennes, le sexe des défunt a été déterminé d'après le mobilier funéraire (**EFFROS 2003 ; 2012 ; LUCY 1997, p. 155**). Dès lors, les sépultures pourvues d'armes sont dites masculines et celles comprenant des éléments de parure sont dites féminines (**HÄRKE 1990 ; EFFROS 2000, p. 632-633 ; KNÜSEL, RIPLEY 2000, p. 157-158**). L'archéologie du genre (gender archaeology) a notamment connu son essor dans les pays scandinaves et anglo-saxons à partir des années 1980 mais a été longtemps boudée par les chercheurs français. Certains travaux récents ont cependant remis cette problématique au cœur des discussions chez les archéologues francophones (**BELARD 2015 ; BELARD 2017 ; SEBILLOTTE-CUCHET 2015 ; BLANCHARD 2020 ; PERIN, BUCHET 2023**). Si aujourd'hui, cette façon d'aborder l'identification des individus est une tradition implicite de la recherche archéologique, elle n'en est pas moins critiquée par les archéologues travaillant sur la notion de genre qui souhaitent observer un renouveau des définitions et des méthodes d'analyse. Leur postulat évoque entre autres le manque de rapprochement avec les données biologiques, mais aussi que le sexe correspond à un caractère physique, lequel ne peut être défini par du mobilier archéologique, qui est un caractère social et qui est le fruit de la main de l'Homme et dont la raison du dépôt dans la tombe nous échappe. La définition du sexe basée sur le mobilier funéraire serait ainsi plutôt simpliste (**BELARD 2017 ; BELARD 2015 ; BLANCHARD 2020 ; KNÜSEL, RIPLEY 2000, p. 159-162 ; LUCY 1997, p. 155**). Les critiques portent également sur la base du raisonnement qui serait lié à des présupposés contemporains et qui impliqueraient que les rôles sociaux liés aux genres seraient les mêmes depuis le V<sup>e</sup> siècle. Par ailleurs, l'absence de « genres alternatifs » seraient dans certains cas le fait d'un déni de leur existence par les historiens plutôt que leur absence effective. (**KNÜSEL RIPLEY 2000 ; p. 158 ; SEBILLOTTE-CUCHET 2015, p. 4**).

Ce travail revêt donc un intérêt multiple. Tout d'abord, il nous permet d'établir l'existence de corrélations entre mobilier funéraire et sexe biologique du défunt, et de proposer ainsi un argument en faveur de leur utilisation. Ensuite, il résulte en une méthode d'annotation automatique exploitant uniquement l'information portée par ce mobilier pour prédire correctement le sexe biologique d'un individu avec 88% de précision. Enfin, à l'aune des deux points précédents, il permet aux chercheurs de choisir d'utiliser le mobilier funéraire comme critère de détermination du sexe biologique, de manière automatique ou non, tout en ayant conscience des biais, ici quantifiés et visualisés, que cela pourrait impliquer.

## Données

### Aperçu

La base de données concernant le mobilier funéraire permet d'obtenir toutes les informations sur un objet, quel qu'il soit. Elle renseigne d'une part la nature de l'objet archéologique initialement déposé dans la tombe en fonction des restes archéologiques qui nous sont parvenus : la découverte d'un umbo permettra par exemple la restitution du dépôt initial d'un bouclier avec le défunt. D'autre part, à l'aide de menus déroulants à choix multiples, le matériau ou les éventuels décors et techniques de fabrication utilisées peuvent être renseignés, ainsi que la datation de l'objet, sa localisation au sein de la structure ou encore sa description détaillée.

### Traitement des données

À chaque sépulture est associée une liste d'objets, un sexe et une tranche d'âge.

Sur la seule base du mobilier funéraire, nous tenterons de distinguer uniquement, dans un premier temps, les individus adultes des individus immatures et les individus masculins des individus féminins.

Du fait de déséquilibres dans la fréquence d'apparition de certaines informations dans le corpus, il est nécessaire de prétraiter les données. Les outils d'apprentissage automatique visent généralement à repérer des régularités statistiques dans les données. Aussi, des informations qui n'apparaissent que très rarement dans le corpus ne permettent aucune analyse pertinente des tendances globales.

Nous avons donc retiré les quelques occurrences de sépultures présentant la classe d'âge « adolescent » ou « adulte-mature », ainsi que les objets apparaissant moins de dix fois au total.

En outre, nous ne considérons pas les objets labélisés comme étant « indéterminés » ou « décontextualisés ». Enfin, les sépultures auxquelles aucun genre et aucun âge n'est associé ne sont pas considérées : l'algorithme ne pourra rien en apprendre, ni s'évaluer dessus.

Au terme de ce prétraitement, le corpus est constitué de 79 sépultures associées à un sexe (48 % féminin, 52 % masculin), et de 334 associées à une classe d'âge (81 % adultes, 19 % immatures). Les premières sont associées à un total de 407 objets, et les secondes à 1668 objets. Un aperçu du corpus final est présenté dans la ([fig. 3](#)).

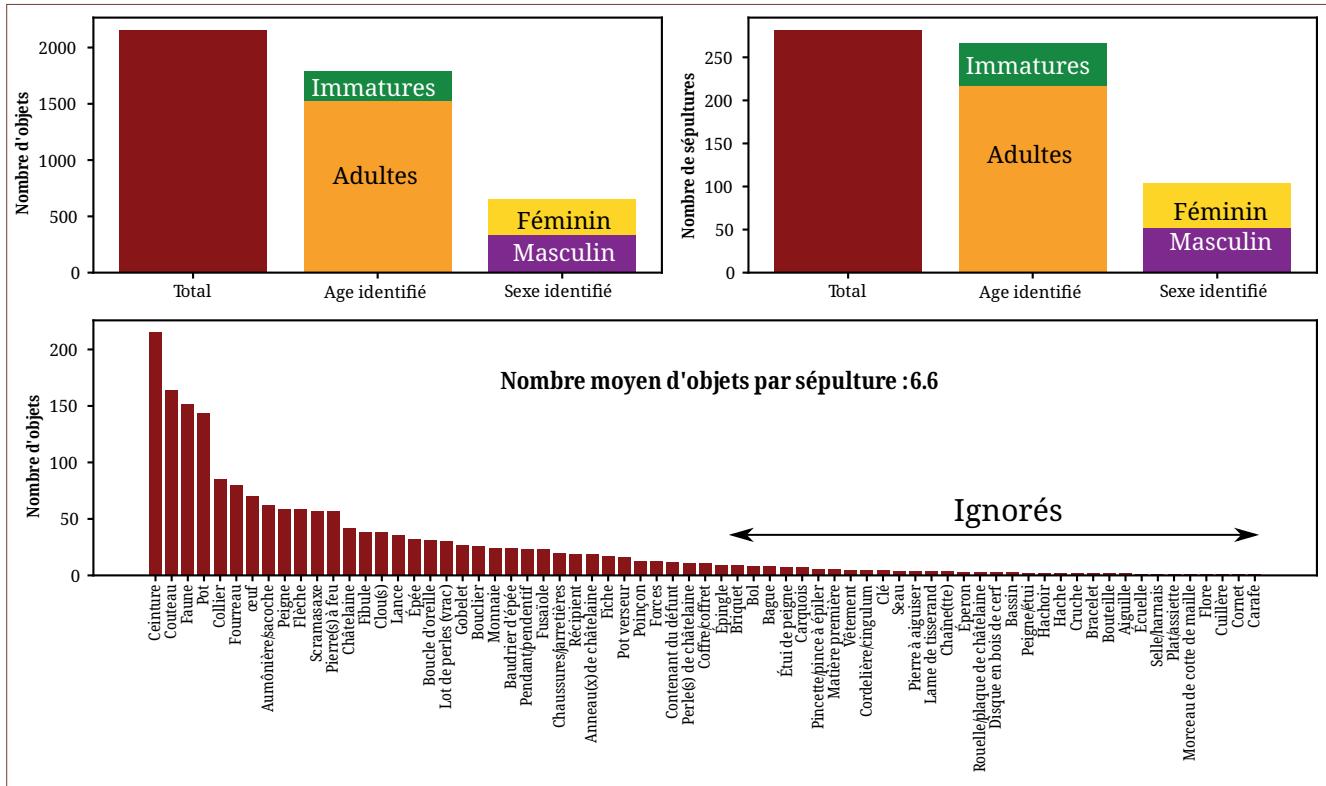


Fig. 3 Corpus considéré par le modèle d'apprentissage automatique.

## Modèle

Cette section vise à expliquer succinctement le fonctionnement du SIMSBM, ce qu'il faut comprendre dans cette notion de clusters et ses limites, inhérentes à tout modèle d'apprentissage automatique.

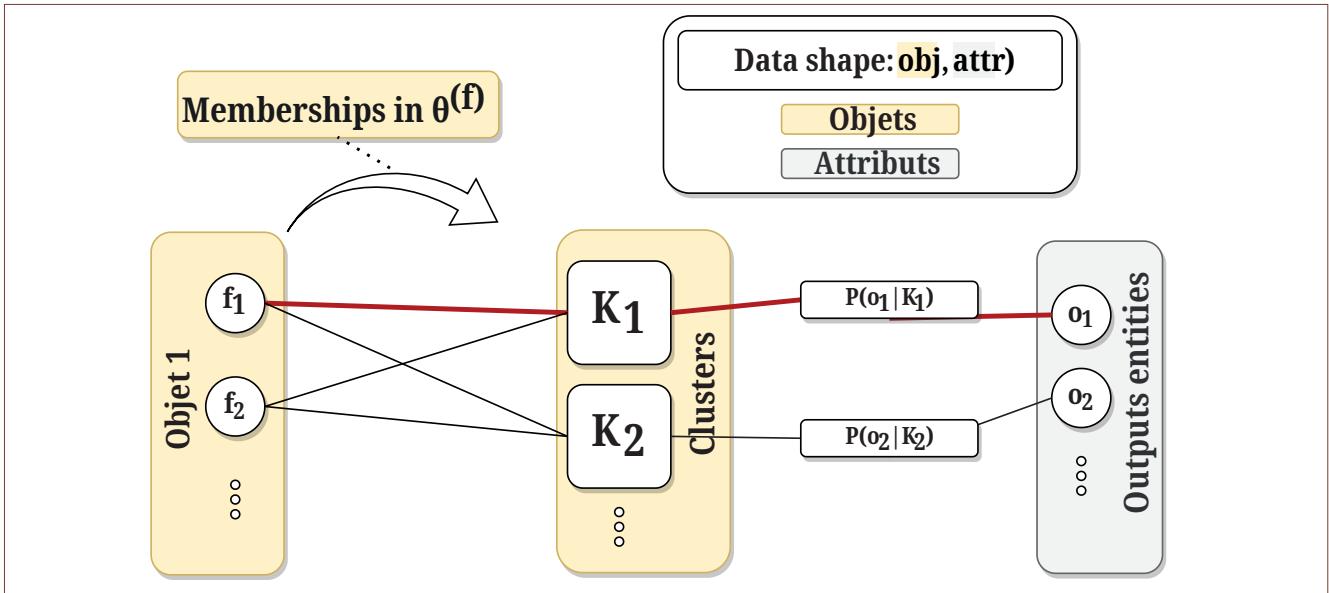
### Principe

Le modèle que nous proposons d'utiliser pour l'analyse de ce jeu de données est nommé SIMSBM, pour Serialized Interacting *Mixed-membership Stochastic Block Model*. Il s'agit d'un modèle associant la probabilité d'un genre ou d'une classe d'âge à chaque objet, paire d'objets ou triplet d'objets. Pour ce faire, il forme automatiquement des groupements d'objets qui présentent des comportements similaires, que nous appellerons clusters. On notera  $P(\text{genre}/\text{âge}) \mid \{\text{objet } 1, \text{objet } 2, \dots\}$  la probabilité d'un genre/âge étant donnés objet 1, objet 2, ...

Le schéma (fig.4) présente la plus simple des versions de SIMSBM, notée SIMSBM(1). Dans ce modèle, chaque catégorie d'objets est répartie entre divers clusters, et chacun de ces clusters est associé à la probabilité d'un attribut (sex ou âge). On note  $\theta_{ik}$  la probabilité que la catégorie d'objets  $i$  appartienne au cluster  $k$ , et on note  $p_k(o)$  la probabilité que le cluster  $k$  soit relié à l'attribut  $o$ .

La probabilité qu'une catégorie d'objets  $i$  soit associée à l'attribut  $o$  est donc égale à la somme de tous les chemins menant  $i$  à  $o$ , c'est-à-dire la somme de  $\theta_{ik} p_k(o)$  pour tous les clusters  $k$ :  $P(i \rightarrow o) = \sum_k \theta_{ik} p_k(o)$

Il s'agit là de la manière dont nous voulons que le modèle représente les données. Nous devons à présent l'entraîner avec des données réelles afin qu'il infère seul la valeur des paramètres  $\theta_{ik}$  et  $p_k(o)$  pour tous les objets  $i$ , les attributs  $o$  et les clusters  $k$ . Enfin, par simplicité, nous représenterons le jeu de données par le symbole D.



**Fig. 4** Schéma du modèle SIMSBM(1) - Les différentes catégories d'objets sont associées dans une certaine proportion  $\theta$  à des clusters. Chacun de ces clusters est à son tour associé avec une certaine probabilité à un attribut, dans notre cas une classe d'âge ou un sexe. Par exemple, le chemin rouge traduit la probabilité que l'objet  $f_1$  soit associé à l'attribut  $o_1$  par le biais du cluster  $K_1$  :  $P(f_1 \rightarrow o_1 | f_1 \in K_1) = P(f_1 \in K_1) \times P(K_1 \rightarrow o_1) = \theta_{f_1 K_1} p_{K_1}(o_1)$ .

## Entraînement

### ❖ Théorème de Bayes

Nous exposons ici le principe général de fonctionnement des algorithmes d'apprentissage automatique dits « Bayésiens », en ce que leur entraînement se fonde sur le théorème de Bayes :

$$P(\theta, p | D) = P(D | \theta, p) \frac{P(\theta, p)}{P(D)}$$

D'après l'équation 1, la probabilité que nos paramètres  $\theta$  et  $p$  (postérieur) décrivent correctement le jeu de données est proportionnelle à la probabilité d'obtenir lesdites données en utilisant des paramètres  $\theta$  et  $p$  donnés (vraisemblance). Les probabilités  $P(\theta, p)$  et  $P(D)$  représentent respectivement la probabilité d'obtenir des paramètres et des données en dehors de tout autre contexte (*a priori*) ; nous ne les considérons pas dans la suite de cet exposé, car nous ne disposons d'aucune information quant aux valeurs *a priori* de ces quantités. Ces termes disparaissent d'eux-mêmes lors de la dérivation mathématique des formules d'apprentissage.

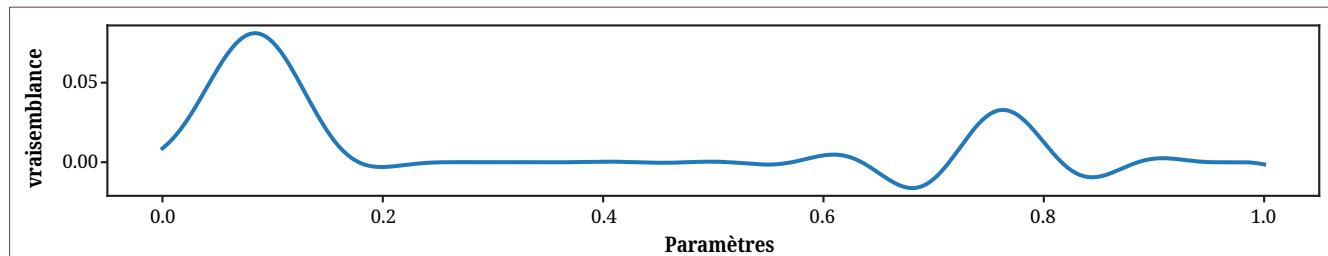
### ❖ Vraisemblance

Le postérieur  $P(\theta, p | D)$  est donc une fonction qui représente la probabilité qu'un jeu de paramètres  $\theta, p$  décrive nos données. Les paramètres qui correspondent à la valeur maximale de cette fonction sont donc ceux qui décrivent le mieux les données. Cependant, cette quantité est impossible à exprimer directement, mais nous voyons dans l'Eq.1 qu'elle est proportionnelle à la vraisemblance  $P(D | \theta, p)$  qui, elle, possède une expression bien définie. Le jeu de paramètres  $\theta, p$  qui maximise cette vraisemblance maximise donc également le postérieur. La vraisemblance s'exprime ainsi :

$$P(D | \theta, p) = \prod_{(i,o) \in D} P(i \rightarrow o) = \prod_{(i,o) \in D} \sum_k^K \theta_{ik} p_k(o)$$

L'équation 2 reste néanmoins difficile à maximiser. Dans la ([fig.5](#)), nous illustrons la difficulté de cette tâche avec un unique paramètre. Dans ce cas, il est évident que le paramètre optimal se trouve autour de 0,1, car nous avons représenté l'intégralité de ces valeurs. Mais en pratique, il est impossible de tracer de telles courbes du fait de trop nombreux paramètres ; leur maximum doit être estimé au moyen d'outils mathématiques. Typiquement, il s'agit de partir d'une valeur aléatoire, d'observer la pente de cette fonction autour de ce point (la dérivée), puis de modifier la valeur de ce paramètre dans le sens de la pente croissante. En itérant ce processus, le paramètre se dirige vers un maximum dit local : lorsque la pente est nulle, la valeur du paramètre ne change plus. Ainsi ([fig. 5](#)), un point de départ qui se situerait autour de 0,7 convergerait vers un maximum local (vers 0,75), mais non global (vers 0,1). Afin d'estimer les meilleurs paramètres pour décrire les données, il est donc nécessaire de relancer cette procédure un grand nombre de fois afin d'optimiser nos chances de trouver un maximum global.

En outre, l'exemple de la ([fig. 5](#)) est délibérément simpliste, en ce que la vraisemblance ne dépend que d'un unique paramètre. Dans notre cas, si nous notons K le nombre de clusters (entre 2 et 10 dans nos expériences), I le nombre de catégories d'objets (35), et O le nombre d'attributs (2), nous pouvons calculer que la vraisemblance à maximiser dépendra de  $(I \times K + K \times O)$  paramètres pour le modèle SIMSBM(1), le plus simple, soit entre 100 et 400 paramètres. Dans le cas du SIMSBM(2) et SIMSBM(3) que nous détaillerons plus tard, il y aura respectivement  $(I \times K + K^2 \times O)$  et  $(I \times K + K^3 \times O)$  paramètres à estimer. Il faut donc s'imaginer que la fonction équivalente à la ([fig.5](#)) doit être représentée dans autant de dimensions qu'il y a de paramètres.



[Fig.5](#) Illustration d'une fonction de vraisemblance pour un seul paramètre. La valeur de ce paramètre qui correspond au maximum de cette courbe est celle qui décrit le mieux le jeu de données.

### ❖ Méthode d'optimisation

Plusieurs démonstrations de méthodes permettant de trouver les maxima locaux de la vraisemblance sont déjà détaillées dans la littérature ([BISHOP 2006](#), [GODOY-LORITE et alii 2016](#), [POUX-MÉDARD et alii 2021 ; 2022](#)). Aussi, nous présentons simplement leurs résultats.

La première étape consiste à calculer la probabilité  $\omega_{i,o}(k)$  qu'un objet i relié à l'attribut o le soit par l'intermédiaire du cluster k, exprimée par :

$$\omega_{i,o}(k) = \frac{p_k(o)\theta_{ik}}{\sum_k^K p_k(o)\theta_{ik}}$$

Cette expression représente la probabilité moyenne qu'un cluster soit utilisé pour relier un objet et un attribut entre eux. Elle est utile en ce qu'elle nous permet de réexprimer l'équation 2 sous une forme équivalente, plus facile à manipuler. Nous rappelons que pour atteindre un maximum local, il faut que les paramètres suivent la pente de la courbe, sa dérivée. En différenciant l'équation 2, nous aboutissons aux expressions suivantes pour la mise à jour des paramètres :

$$\theta_{mk} = \frac{\sum_{(i,o) \in R^o | i=m} \omega_{io}(k)}{\sum_{(i,o) \in R^o | i=m} 1}$$

$$p_k(n) = \frac{\sum_{(i,o) \in R^o | o=n} \omega_{io}(k)}{\sum_{(i,o) \in R^o} \omega_{io}(k)}$$

La procédure consiste donc à alterner entre le calcul explicite de  $\omega_{io}(k)$  à l'aide de l'équation 3 et le calcul des nouveaux paramètres  $\theta$  et  $p$  à l'aide des équations 4, ce jusqu'à convergence vers un maximum local. À l'aboutissement de cette procédure, nous obtenons donc les valeurs numériques pour chacune des entrées  $\theta_{ik}$  et  $p_k(o)$  qui correspondent aux paramètres localement optimaux pour décrire le jeu de données.

## Variantes de SIMSBM

Dans nos expériences, nous testerons trois variations de SIMSBM, dont la représentation schématique est donnée dans la ([fig. 6](#)). Dans un premier temps, considérons un exemple simple. Le SIMSBM s'entraîne sur les 3 sépultures suivantes :

- S1 : pince, épée, genre masculin
- S2 : pince, épingle, genre féminin
- S3 : chaînette, genre féminin

Le modèle le plus simple va chercher à associer un objet à un attribut. Il s'entraînera en considérant (pince, masculin), (épée, masculin), (pince, féminin), (épée, féminin) et (chaînette, féminin). En sortie, il fournira la probabilité d'un attribut étant donné un objet uniquement,  $P(\text{féminin} | \text{pince})$ , par exemple. Ce modèle est dénoté par SIMSBM(1) —([fig. 6-haut](#))

Un modèle plus compliqué considérera des paires d'objets. Il associera chaque paire à la probabilité d'un attribut. Dans notre exemple, il s'entraînera sur (pince, pince, masculin), (pince, épée, masculin), (épée, épée, masculin), (pince, pince, féminin), (pince, épingle, féminin), (épingle, épingle, féminin) et (chaînette, chaînette, féminin). En sortie, il pourra fournir des probabilités comme  $P(\text{masculin} | \text{pince, épingle})$ . Ce modèle est dénoté par SIMSBM(2) —([fig. 6-milieu](#))

Un modèle encore plus compliqué considérera des triplets d'objets en entrée ; on le nomme SIMSBM(3). Il pourra fournir des probabilités comme  $P(\text{féminin} | \text{pince, épingle, épingle})$  — ([fig. 6-bas](#))

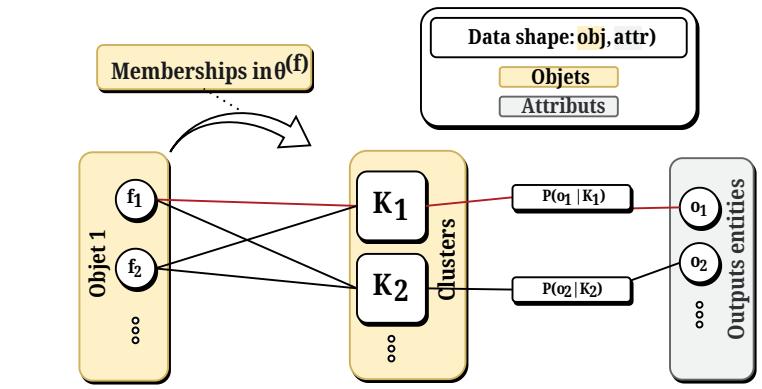
Plus le nombre d'objets en entrée est grand, plus le modèle peut être exhaustif dans sa description. On se doute que dans la plupart des cas réels, plus de trois objets sont retrouvés dans une sépulture. Dans ces cas, une fois le modèle entraîné, la prédiction finale sera la moyenne de la prédiction pour chaque combinaison d'objets. Cependant, compte tenu du nombre de données à notre disposition, il est peu probable que le modèle parvienne à trouver des corrélations significatives entre plus de trois objets à la fois, car très peu de combinaisons d'un tel ordre se répètent dans le corpus. Pour cette raison, nous n'utiliserons pas SIMSBM(4) ou plus.

## Groupes

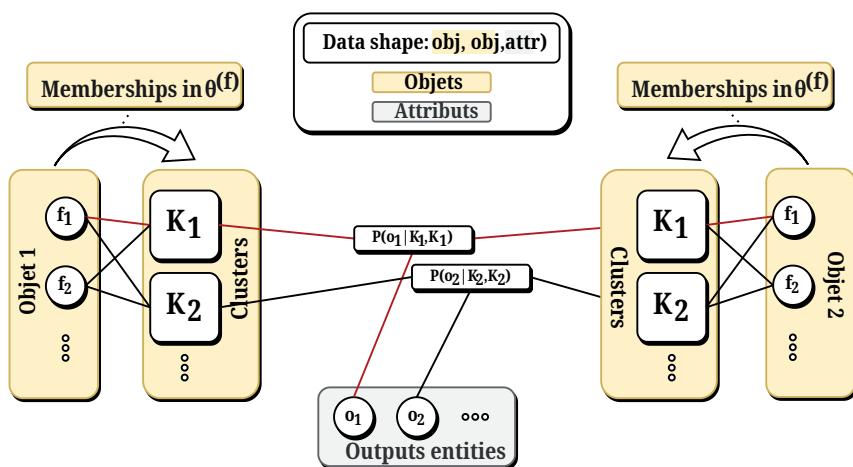
Pour aboutir à ces prédictions, SIMSBM fonctionne en classant les objets dans des groupes (ou clusters), puis en inférant les probabilités qu'ont ces groupes d'être reliés à un attribut donné. Cela signifie que les objets ne sont pas directement associés à un attribut, mais seulement à un cluster, qui lui y sera associé.

Le nombre de groupes est fixé à l'avance, et doit être inférieur au nombre d'objets. Cela permet de forcer le modèle à ranger intelligemment les objets dans les différents groupes pour fournir des prédictions solides — formulé autrement, à trouver des régularités dans les données.

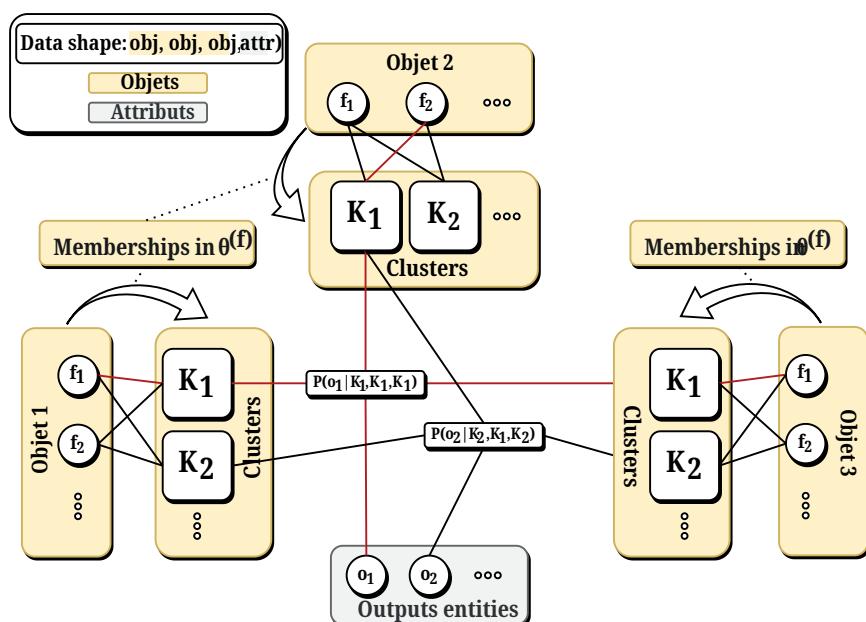
## SIMSBM(1)



## SIMSBM(2)



## SIMSBM(3)



**Fig. 6** Différentes itérations du modèle SIMSBM que nous utiliserons dans nos expériences

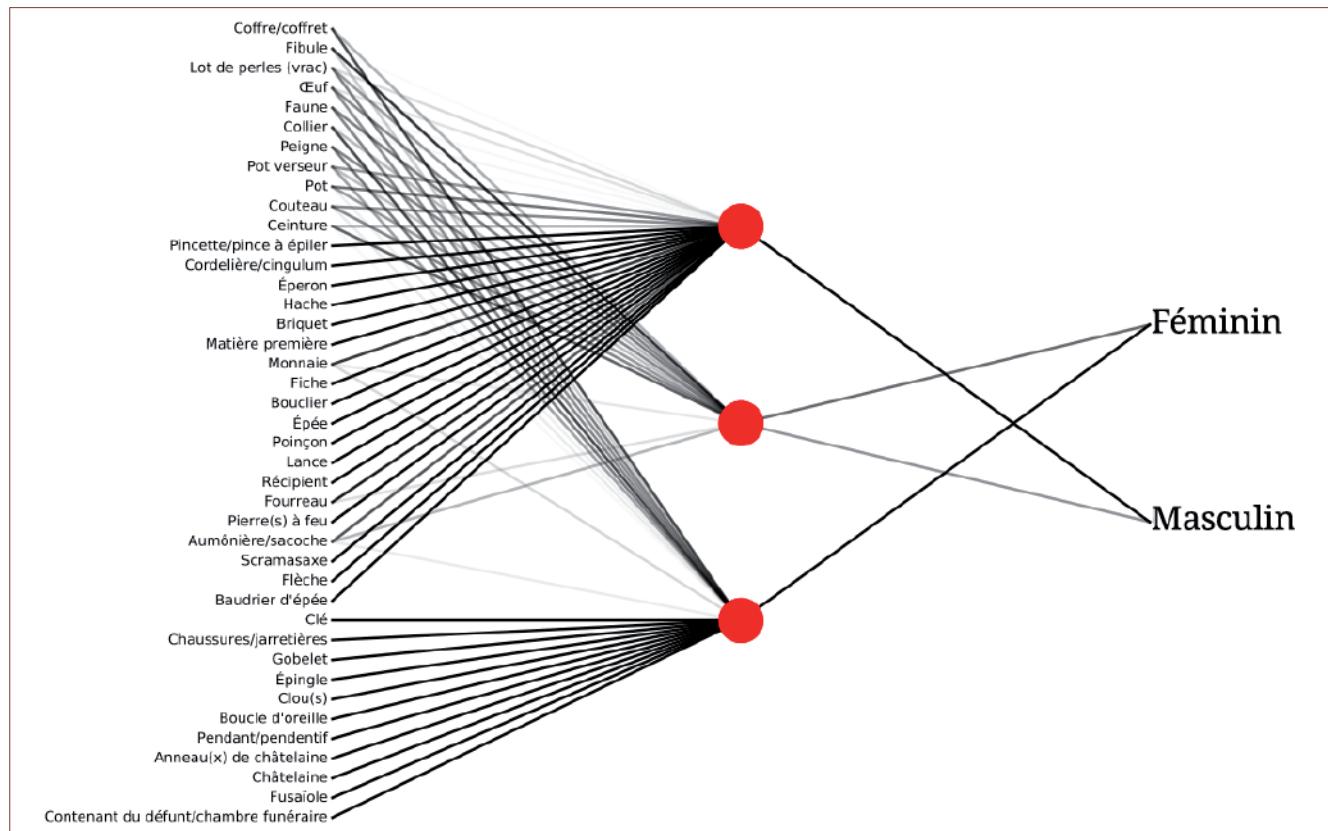
Nous illustrons cela dans la ([fig. 7](#)). Les objets appartiennent plus ou moins à chacun des trois groupes et les groupes sont reliés avec une probabilité plus ou moins grande à chacun des genres possibles, comme vu précédemment. Ainsi, le groupe du haut est le groupe des objets à caractère masculin : les objets appartenant fortement à ce groupe (la hache par exemple) ont donc de fortes chances d'être associés à un individu masculin. Le groupe du milieu semble fournir une probabilité égale pour chacun des genres ; les objets y appartenant fortement (la fibule dans cet exemple) ne sont donc pas déterminants dans l'estimation du genre. Enfin, le groupe du bas est fortement féminin et les objets qui y appartiennent ont donc de fortes chances d'être associés à un individu de ce genre. D'autres objets, comme l'œuf, appartiennent de manière significative à deux groupes à la fois (ici le groupe 2 « féminin » et le groupe 3 « indécis ») et sont donc des objets dont l'association à un genre donné est plus floue.

Les autres itérations de SIMSBM considérant un plus grand nombre d'objets en entrée présentent une structure similaire. Pour SIMSBM(2), à la place de groupes uniques ([milieu de fig. 7](#)) se trouveront des couples de groupes. Chaque objet sera toujours lié à plusieurs groupes, mais ces groupes seront reliés entre eux en paires avant de fournir leur probabilité. SIMSBM(3) suit la même logique, où chaque objet appartient à plusieurs groupes, et chaque triplet de groupes fournira une probabilité pour chaque genre.

## Limites

- Lors de l'entraînement du modèle, les paramètres convergent vers un maximum de vraisemblance *local*. Relancer deux fois l'apprentissage sur un même jeu de données fournira ainsi des résultats légèrement différents en termes d'organisation des groupes. Il faut donc entraîner le modèle un grand nombre de fois (typiquement 100) afin d'explorer au maximum l'espace des paramètres. Le modèle le plus vraisemblable, celui pour lequel l'équation 2 a la plus grande valeur, est sélectionné au terme de ces 100 entraînements.
- Le nombre de clusters n'est pas inféré automatiquement. Afin de le déterminer, nous l'entraînerons plusieurs fois pour un nombre de clusters variant de 2 à 10. Nous sélectionnerons ensuite le nombre de groupes à partir duquel les performances du modèle n'augmentent plus.
- Les groupes sont rarement aussi facilement interprétables qu'en ([fig. 7](#)). Il faut systématiquement conduire des analyses particulières des groupes trouvés et de leur contenu pour deviner ce qu'ils représentent. Imaginons que le modèle décèle que les objets en fer sont masculins dans 70 % des cas et qu'il crée un groupe spécifique à ces objets avec 70 % de probabilité vers le masculin et 30 % vers le féminin : il faudra attentivement analyser la sortie du modèle pour comprendre ce qu'il aura déjà compris.
- Rejoignant le point suivant, il ne faut pas perdre de vue que les modèles d'apprentissage automatique n'effectuent aucun raisonnement, mais cherchent simplement à décrire le plus vraisemblablement des données qui leur sont fournies. Ils ne possèdent aucune information ni sur ce qu'est une fibule, ni sur sa fonction ; il s'agit simplement là d'une étiquette, qui apparaîtra plus ou moins régulièrement avec d'autres étiquettes, et toutes celles-là seront associées à des attributs. SIMSBM et les modèles d'apprentissage automatique, doivent être considérés comme les outils qu'ils sont. Ils ne peuvent en aucun cas se substituer à l'analyse d'un chercheur, visant plutôt à l'assister dans ses études.

- Les prédictions du modèle, on l'a dit, sont faites sur des triplets, paires ou singlets d'objets. Or, le nombre d'objets par sépulture peut varier. Afin d'obtenir les probabilités de tel ou tel attribut associé, notre méthode consiste à faire la moyenne de toutes les probabilités qui y sont associées. Si une sépulture contient un peigne et une épée, et que les probabilités inférées indiquent  $P(\text{femme} | \text{épée}) = 0,3$  et  $P(\text{femme} | \text{peigne}) = 0,5$ , la probabilité que la sépulture soit celle d'une femme vaudra 0,4. Cette méthode fournit de bons résultats, mais induit un biais expérimental post-entraînement. D'autres approches d'agrégation de ces probabilités (valeur médiane, probabilité maximale, etc.) seraient également valides.



**Fig. 7** SIMSBM(1) pour 3 groupes. Le graphe se lit de gauche à droite : les objets appartiennent à l'un des trois groupes dans une proportion indiquée par la couleur des traits les y reliant. Les groupes fournissent ensuite la probabilité de chacun des genres.

## Expériences

### Procédure

L'évaluation consiste en la prédiction des attributs d'une sépulture uniquement sur la base du petit mobilier qui y a été retrouvé. Les paramètres inférés par les modèles doivent présenter une valeur générale, qui s'étend au-delà des données sur lesquelles ils se sont entraînés. Afin d'évaluer leur pouvoir prédictif, nous séparons donc le jeu de données en deux : une partie constituée de 80 % des observations sera utilisée pour l'entraînement du modèle, tandis que les 20 % restants serviront à son évaluation. Ainsi, le modèle devra deviner les attributs de sépultures qu'il n'a jamais observées, ce qui permet d'apprécier la généralité de ses résultats.

Cependant, un mauvais découpage des données peut conduire à des biais. Si, par hasard, les 20 % de données d'évaluation ne comprenaient que des tombes masculines, un modèle qui ne prédirait que « masculin » aurait 100 % de réussite, sans pour autant que cela prouve son efficacité. Afin d'éviter cet écueil, nous utilisons une procédure

de validation croisée, représentée dans la (fig. 8). Le modèle s'entraîne cinq fois sur 80 % des données, de telle sorte que les 20 % de données réservées à l'évaluation ne comprennent jamais les mêmes observations. Ainsi, en nous évaluant indépendamment cinq fois sur cinq jeux de données différents, nous éliminons de possibles biais introduits par ce découpage. Le résultat reporté sera la moyenne de la précision obtenue sur chacun de ces cinq jeux d'évaluation.

Une fois le modèle entraîné, nous procédons à l'évaluation comme suit : pour chaque singlet, paire ou triplet d'objets dans les sépultures du jeu de données test, nous demandons au modèle de nous fournir les probabilités d'attributs correspondantes. Nous prenons la moyenne de ces probabilités, que nous interprétons comme la probabilité qu'une tombe possède ces attributs. La plus haute de ces probabilités déterminera la prédiction du modèle. Une sépulture qui aurait 52 % de chances d'être masculine et 48 % d'être féminine, sera prédictée comme « Masculine » lors de l'évaluation. Enfin, nous reporterons nos résultats en termes de précision, c'est-à-dire le nombre de prédictions correctes divisé par le nombre total de prédictions.



**Fig. 8** Validation croisée – Le modèle s'entraîne cinq fois sur 80% des données, de telle sorte que les 20% de données réservées à l'évaluation ne comprennent jamais les mêmes observations. Ainsi, en nous évaluant indépendamment cinq fois sur cinq jeux de données différents, nous éliminons de possibles biais introduits par ce découpage.

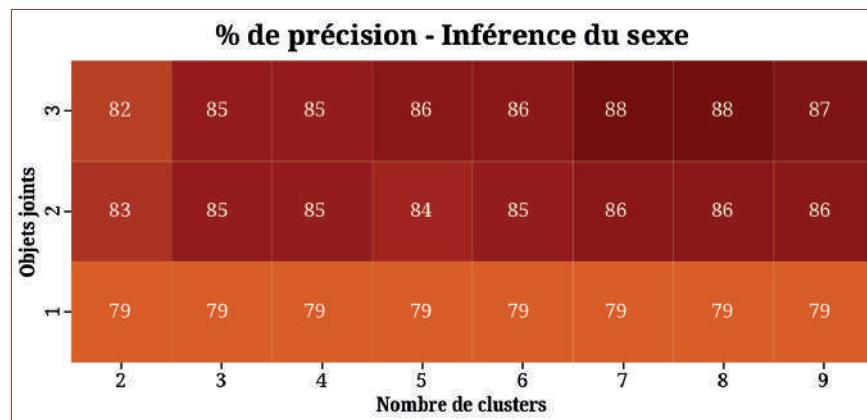
## Résultats pour l'inférence de genre

### ❖ Résultats numériques

Dans la (fig. 9), nous représentons les performances de SIMSBM(1), SIMSBM(2) et SIMSBM(3) dans une tâche d'inférence du genre des individus, pour un nombre de clusters variant de 2 à 10. Les valeurs reportées sont la moyenne des précisions obtenues après validation croisée. Nous rappelons que notre corpus comprend environ 50 % d'individus masculins et 50 % d'individus féminins, si bien qu'un mauvais modèle ferait des prédictions justes 50 % du temps en moyenne.

Nous voyons que le modèle le plus performant est SIMSBM(3) avec 7 clusters. À partir du seul mobilier retrouvé dans les sépultures, il parvient à prédire le sexe correct dans 88 % des cas. Cette constatation implique qu'il existe de fortes corrélations entre petit mobilier et sexe des individus et que SIMSBM parvient à les identifier. Nous notons également que les résultats de SIMSBM(2) et SIMSBM(1) restent corrects : des corrélations existent entre des objets seuls ou des paires d'objets et le sexe des individus inhumés, bien que celles-ci soient vraisemblablement moins déterminantes qu'avec des triplets d'objets.

Outre la mise en évidence de telles corrélations, ces résultats permettent d'entrevoir un usage pratique de ces résultats. Nous rappelons que dans notre jeu de données de base, seules 80 sépultures sur les 2500 mises à notre disposition étaient associées à un sexe. À l'aune de ces résultats, il serait envisageable d'enrichir les données d'une nouvelle méthode d'identification du sexe fondée sur l'étude du mobilier, dont nous saurions qu'elle est vraisemblablement correcte dans 88 % des cas, et permettre ainsi des études ultérieures. Une telle caractérisation n'est évidemment pas à l'ordre du jour pour plusieurs raisons : 12 % d'erreur est une marge bien trop grande, le modèle ne s'est entraîné que sur peu de tombes, probablement peu représentatives de l'ensemble du corpus, et ce pour une période spécifique. Cependant, ces écueils disparaîtront à mesure que la quantité de données d'entraînement augmentera, ce qui se profile avec l'apparition de bases de données archéologiques massives, pour peut-être permettre des caractérisations se rapprochant des 100 % de précision. Rappelons qu'en anthropologie biologique, il est aujourd'hui, du moins en France, classiquement admis par l'ensemble de la communauté scientifique que le seuil de fiabilité retenu pour une méthode de détermination du sexe doit être supérieur ou égal à 95%.



**Fig. 9** Résultats en prédiction du sexe en fonction du nombre de clusters et du nombre d'objets considérés lors de l'entraînement (en % de précision).

mobilier, dont nous saurions qu'elle est vraisemblablement correcte dans 88 % des cas, et permettre ainsi des études ultérieures. Une telle caractérisation n'est évidemment pas à l'ordre du jour pour plusieurs raisons : 12 % d'erreur est une marge bien trop grande, le modèle ne s'est entraîné que sur peu de tombes, probablement peu représentatives de l'ensemble du corpus, et ce pour une période spécifique. Cependant, ces écueils disparaîtront à mesure que la quantité de données d'entraînement augmentera, ce qui se profile avec l'apparition de bases de données archéologiques massives, pour peut-être permettre des caractérisations se rapprochant des 100 % de précision. Rappelons qu'en anthropologie biologique, il est aujourd'hui, du moins en France, classiquement admis par l'ensemble de la communauté scientifique que le seuil de fiabilité retenu pour une méthode de détermination du sexe doit être supérieur ou égal à 95%.

#### ❖ Interprétation

Comme nous l'avons souligné lorsque nous avons évoqué les limites des modèles d'apprentissage automatique, l'objectif de ces approches reste avant tout d'assister une recherche qui ne se limite pas à des chiffres et des prédictions. Un modèle peut prédire le sexe des individus dans 90 % des cas, et ensuite ? Comment aboutit-il à ces prédictions, que peut-on en tirer, quelles corrélations a-t-il identifiées au milieu de cette masse de données à première vue informe ? Pour répondre à ces questions, nous proposons dans cette section plusieurs visualisations permettant aux chercheurs concernés d'interpréter ces résultats.

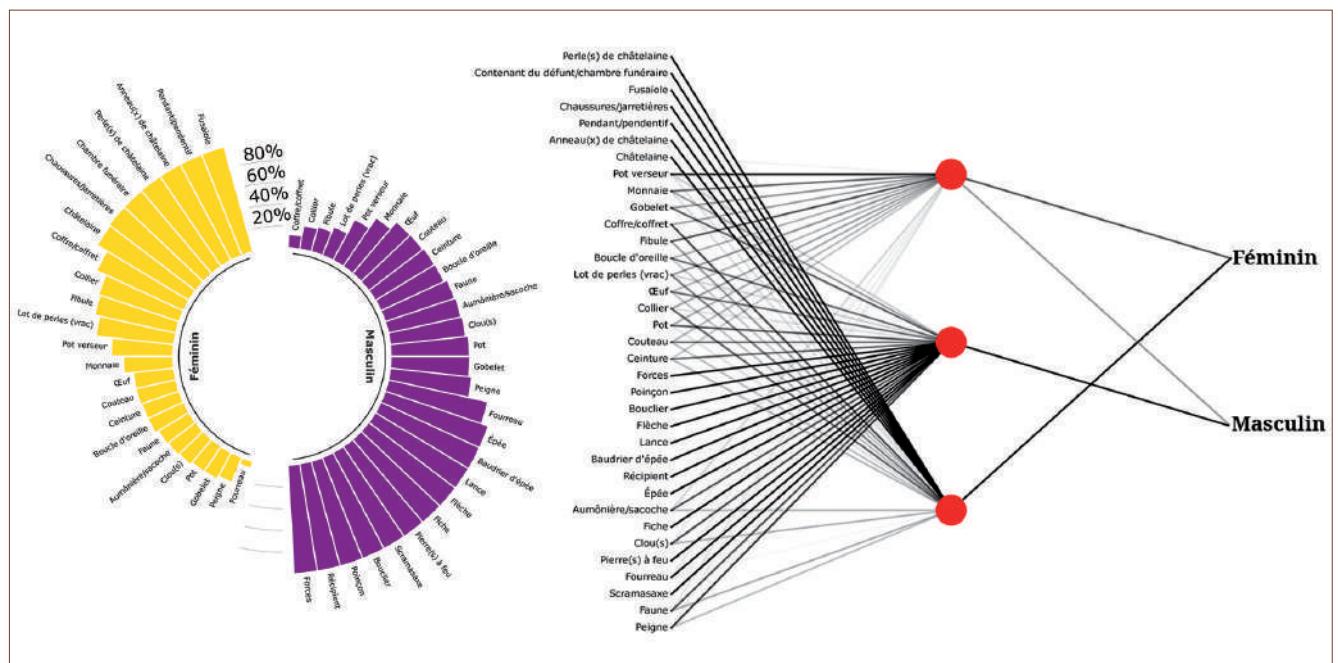
Dans la (**fig. 10**), nous présentons le modèle SIMSBM(1) à trois clusters après entraînement (79 % de précision). Nous voyons à gauche que le modèle a inféré un cluster relié exclusivement au sexe masculin, un autre au sexe féminin, et un autre également relié aux deux (groupe indéterminé). Un objet comme les « perles de châtelaines » sont donc des objets associés à une sépulture féminine dans 100 % des cas, d'après le modèle. D'autres, comme les « boucliers », sont 100 % masculins. D'autres, enfin, comme les « peignes », sont associés tantôt aux sépultures féminines, tantôt aux masculines. Ces associations se retrouvent dans la figure de droite, qui représente la même information sous une forme plus synthétique. Nous observons donc là les paramètres d'un modèle qui n'a raison que dans 79 % des cas.

Dans la (**fig. 11**), nous représentons la même figure que la (**fig.10-droite**) pour SIMSBM(2) avec 7 clusters. Pour rappel, les prédictions se font à présent à partir de paires d'objets, plutôt que d'objets seuls. Aussi, la couleur de chaque case de la (**fig.11-gauche**) représente la probabilité qu'un couple d'objets soit associé à un sexe masculin ou féminin. Trois zones se dégagent de cette figure. Tout d'abord, une zone dans laquelle tous les couples d'objets sont associés à une sépulture masculine avec 100 % de probabilité. C'est le cas, par exemple, de la paire « Lance »

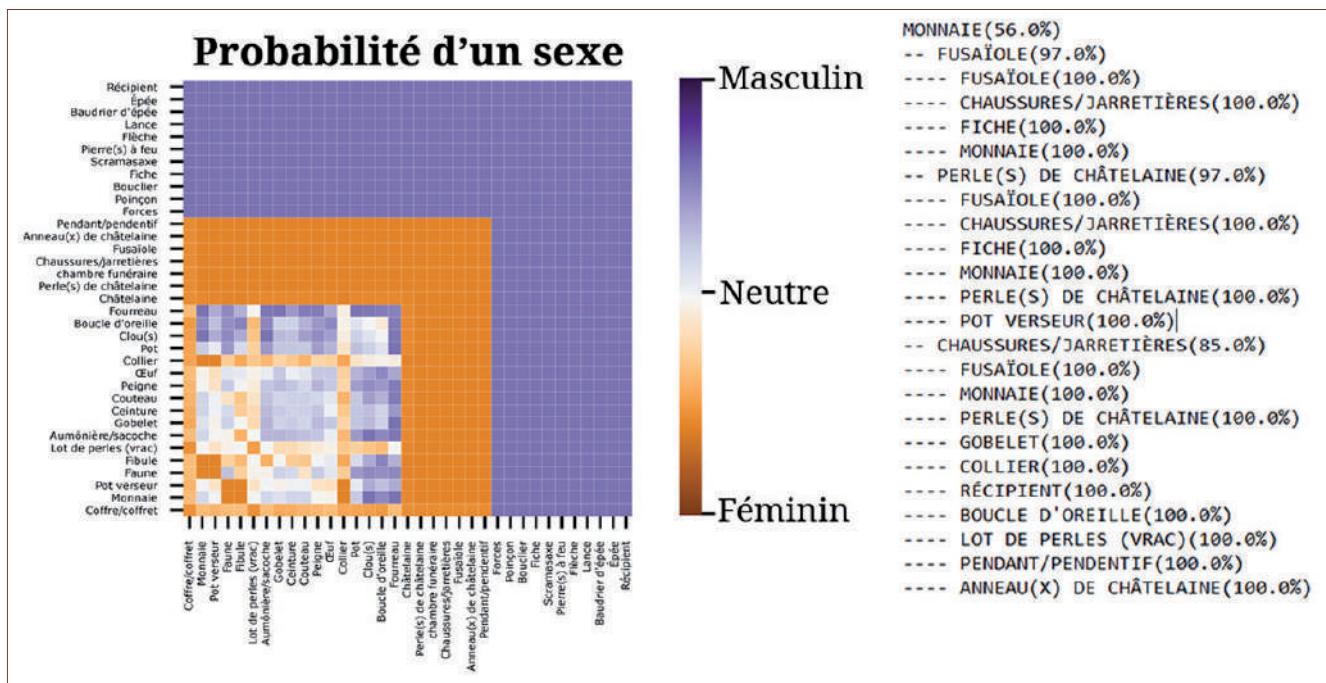
et « Scramasaxe ». Une seconde zone, en orange, représente tous les couples associés exclusivement aux sépultures féminines (paire « Fusaïole » et « Pendentif », par exemple). Une troisième zone, enfin, plus floue, où le sexe inféré dépend fortement de la paire d'objets considérés. Un dépôt de « Monnaie » associé à un « Collier » implique une grande probabilité de sépulture féminine, mais les dépôts de « Monnaie » et de « Clous » sont plutôt associés aux sépultures masculines.

De la même manière, il est intéressant de remarquer que certains objets dépendent fortement de la seconde partie de leur paire pour fournir une prédiction. Par exemple, la fusaïole est associée avec 100 % de chances à des tombes féminines lorsqu'elle se trouve accompagnée d'un pendentif, d'anneaux de châtelaine ou de chaussures, mais est associée avec 100 % de chances à une tombe masculine lorsqu'associée à un scramasaxe, une flèche, etc. Cela peut être dû à une véritable corrélation, où un objet prendrait le pas sur l'autre, ou, plus vraisemblablement, à un biais dans la méthode de visualisation. Les probabilités représentées ici ne prennent pas en compte le nombre de données qu'a reçu le modèle lors de l'entraînement, et il est fort possible que la paire « fusaïole + scramasaxe » n'ait pas été observée. Dans ce cas, le modèle convergerait vers la valeur de l'un ou de l'autre, sans distinction, car aucune information ne lui permettrait d'en apprendre quoi que ce soit. Cet aspect doit être pris en compte lors de l'exploitation de ces résultats. Il n'en reste pas moins que c'est bien le système représenté dans la ([fig.11-gauche](#)) qui fournit des prédictions justes dans 86 % des cas et identifie plusieurs corrélations claires entre paires de petit mobilier et sexe des individus.

Enfin, concernant SIMSBM(3), nous n'avons pas trouvé de visualisation graphique satisfaisante — il faudrait représenter un cube plein, où chaque zone représenterait la probabilité d'un sexe étant donné un triplet d'informations. Le modèle ne peut s'analyser dans le détail qu'en représentant chacune des 35 tranches qui le composent, soit sous forme de liste à niveaux, comme celle représentée dans la ([fig. 11-droite](#)). Malgré la taille de la liste (nombre d'objets au cube), il est possible de l'exploiter par des moyens informatiques simples.



**Fig. 10** Visualisations du lien entre petit mobilier et sexe pour SIMSBM(1) – À gauche, probabilité qu'un objet appartienne à un cluster et qu'un cluster soit relié à un sexe. La couleur des liens représente leur force ; un lien noir représente 100 % de probabilité, et un lien gris pâle/blanc représente 0 % de probabilité. À droite, pour le même modèle, probabilité qu'un objet seul soit associé à un sexe en particulier. Probabilités obtenues en sommant tous les chemins menant d'un objet à un sexe.



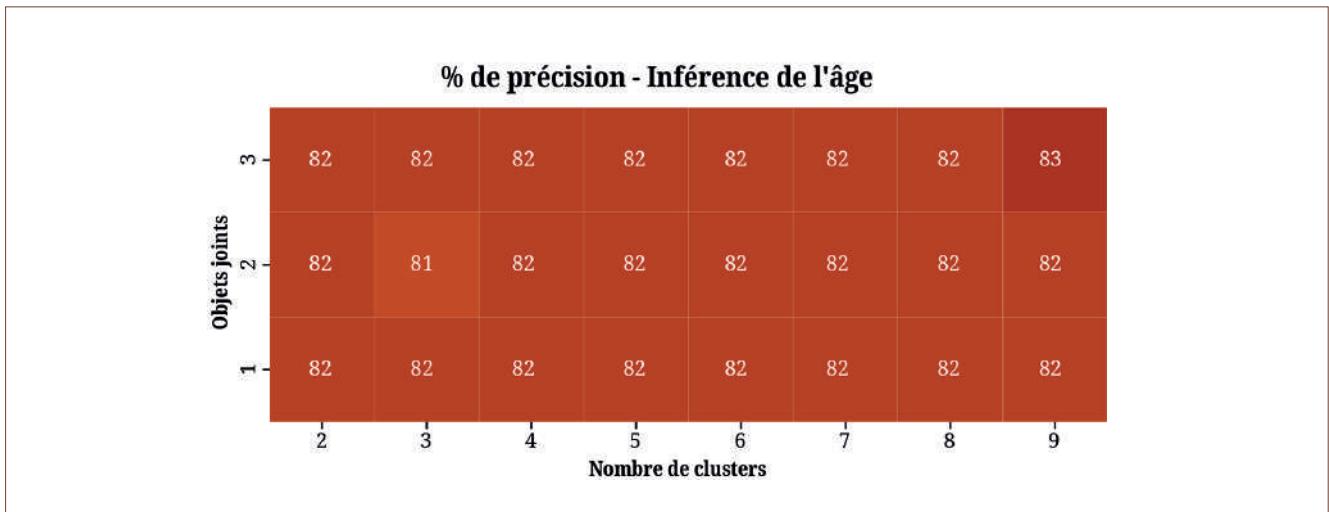
**Fig. 11** À gauche, visualisation du lien entre petit mobilier et sexe pour SIMSBM(2) – Probabilité qu'une paire d'objets soit associée à un sexe en particulier. Probabilités obtenues en sommant tous les chemins menant d'une paire d'objets à un sexe. À droite, visualisation pour SIMSBM(3) — Les pourcentages correspondent à la probabilité que l'individu associé soit une femme (56 % s'il s'agit d'une monnaie seule, 85 % si elle se trouve avec des chaussures, et 100 % s'il s'y trouve également une fusaïole).

## Résultats pour l'inférence de l'âge

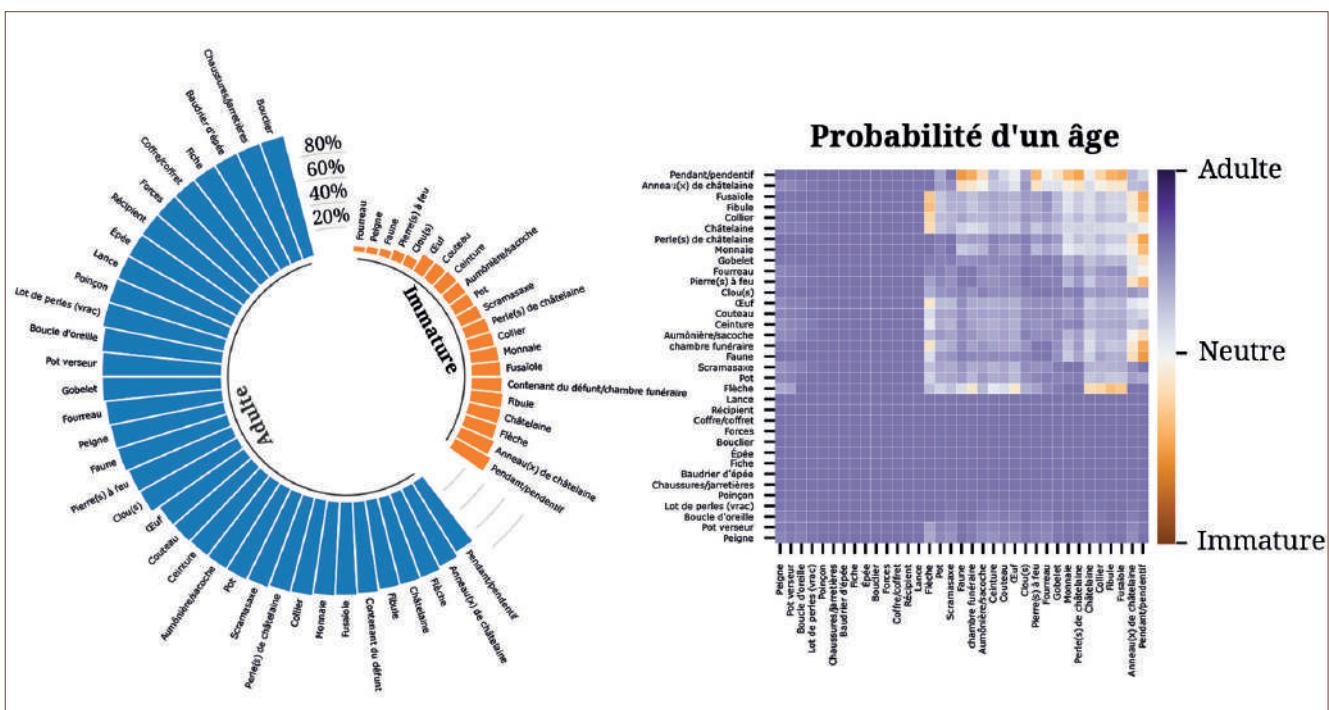
Dans la (fig. 12), nous représentons les performances de SIMSBM(1), SIMSBM(2) et SIMSBM(3) dans une tâche d'inférence du genre des individus, pour un nombre de clusters variant de 2 à 10. Les valeurs reportées sont la moyenne des précisions obtenues après validation croisée. Nous rappelons que notre corpus comprend environ 81 % d'individus adultes et 19 % d'individus immatures, si bien qu'un mauvais modèle ferait des prédictions justes 81 % du temps en moyenne.

Contrairement à ce qui a été observé pour le sexe, aucune des versions de SIMSBM ne parvient à fournir de résultats exploitables. Il peut y avoir deux explications à cet échec : soit le modèle ne possède pas assez de données pour identifier des corrélations significatives entre petit mobilier et classe d'âge, soit cette corrélation n'existe simplement pas pour les classes d'âges définies. Dans le premier cas, il s'agirait du déséquilibre entre le pourcentage d'individus adultes et celui d'individus immatures ; les observations ne permettent globalement pas de bien les différencier, et les modèles convergeraient vers des *maxima* locaux de la fonction de vraisemblance, qui les pousseraient à ne prédire que la classe « adulte ». Dans le second cas, soit il n'existe tout simplement pas de corrélation entre petit mobilier et classe d'âge, soit les classes d'âges telles que définies ne permettent pas cette analyse. Certains types de dépôts pourraient être partagés entre les plus jeunes adultes et les plus vieux immatures, rendant les deux classes impossibles à distinguer. En outre, des variations pourraient exister à l'intérieur même de la classe « adultes », sans que le modèle puisse les repérer en l'absence d'informations supplémentaires.

Ces constatations peuvent être observées dans la (fig. 13), où nous constatons que l'écrasante majorité des objets est associée à la classe « adultes » au détriment de la classe « immature ». De ce graphique, nous ne pouvons que constater la nécessité d'accentuer les efforts de mise en commun des bases de données archéologiques massives afin de permettre, soit de trancher la question et d'établir la non-existence de corrélations entre petit mobilier et classe d'âge, soit d'observer des corrélations manifestement faibles.



**Fig.12** Résultats en prédiction de la classe d'âge en fonction du nombre de clusters et du nombre d'objets considérés lors de l'entraînement (en % de précision).



**Fig. 13** Visualisations du lien entre petit mobilier et classe d'âge pour SIMSBM(1) et SIMSBM(2) – Probabilité qu'un objet seul (gauche) ou une paire d'objets (droite) soit associé à une classe d'âge en particulier.

## Conclusion

Dans cet article, nous avons proposé une méthode d'apprentissage automatique afin d'exploiter un corpus conséquent provenant de récents efforts de mise en commun des données scientifiques. Nous avons commencé par détailler la méthode proposée, SIMSBM, puis expliqué les grandes lignes de la procédure d'entraînement de ce modèle. Son application à notre corpus, une liste de sépultures mérovingiennes associées au sexe et à la classe d'âge de leur occupant, ainsi qu'aux objets retrouvés avec lui, nous a permis d'établir l'existence de corrélations entre ces éléments.

Nous avons ainsi montré la possibilité d'inférer le sexe d'un individu sur la base exclusive du mobilier funéraire, avec une précision de 88 %. Ce résultat indique à lui seul l'existence de pratiques funéraires distinctes en fonction du sexe du défunt. En outre, les quelques 12 % de marge d'erreur du modèle, vraisemblablement amenés à diminuer avec l'apport de données nouvelles, laissent entrevoir la possibilité, à terme, d'utiliser le SIMSBM pour des tâches d'annotation automatisée de corpus archéologiques.

Nous avons ensuite montré que le SIMSBM ne parvenait à détecter aucune corrélation entre mobilier funéraire et classe d'âge dans ce corpus. Que ce soit du fait de leur non-existence, ou de données trop peu abondantes pour les dévoiler, cette constatation nous a permis de souligner la nécessité de continuer les efforts de mise en commun et de structuration des données de recherche, afin de pouvoir statuer de manière claire sur des cas plus ambigus, comme celui-ci.

Enfin, cet article vise plus généralement à mettre en évidence les bénéfices d'une recherche pluridisciplinaire. L'émergence de bases de données archéologiques massives (Artefacts, Clauss-Slaby, et différents PCR portant sur le Grand Est, mais aussi l'Île-de-France ou la Champagne<sup>4</sup>) s'accompagne de la nécessité de développer des outils de traitement et d'exploitation automatique de l'information. Les méthodes d'apprentissage automatique se sont longtemps confrontées à des difficultés d'application au champ des sciences humaines, du fait d'une insuffisance en données disponibles et utilisables numériquement. Cette tendance change à présent, ouvrant la communauté archéologique à une panoplie de nouveaux outils, et autant de résultats nouveaux découlant de leur application.

---

4 Cf. actes du présent colloque, AFAM XLII

## Bibliographie

**BARRAND EMAM H., CHENAL F., ABERT F.** 2022  
*Espaces et pratiques funéraires en Alsace aux époques mérovingienne et carolingienne (V<sup>e</sup>-X<sup>e</sup> siècles)*, Projet Collectif de Recherche, 2015-2022, Rapport d'activité 2021, SRA Alsace, février 2022, 187 p.

**BARRAND EMAM et al.** à paraître  
*Le Projet collectif de recherche « Espaces et pratiques funéraires en Alsace aux époques mérovingienne et carolingienne (V<sup>e</sup>-X<sup>e</sup> siècles). Bilan critique de six années de recherches*, Actes du colloque de l'AFAM, Saint-Germain-en-Laye, octobre 2022, à paraître.

**BELARD C.** 2015  
La notion de genre ou comment problématiser l'archéologie funéraire, *Les Nouvelles de l'Archéologie*, 140, 2015, p.23-27

**BELARD C.** 2017  
*Pour une archéologie du genre : Les femmes en Champagne à l'âge du fer*. Hermann, 272p.

**BISHOP C.** 2006  
*Pattern Recognition and Machine Learning*. Springer-Verlag New York 2006, 738p.

**BLANCHARD C.** 2020  
« Pratiques funéraires, mobilier funéraire et genre : état des lieux des études mérovingiennes dans le bassin parisien » dans ALGRAIN I. (dir.), *Archéologie du genre : construction sociale des identités et culture matérielle*, Bruxelles, Université des Femmes.

**BRUZEK J., SANTOS F., DUTAILLY B., MURAIL P., CUNHA E.** 2017

Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology, *American Journal of Physical Anthropology*, 164, 2, p. 440-449.

**EFFROS B.** 2000  
Skeletal sex and gender in Merovingian mortuary archaeology, *Antiquity*, 74-285, 2000, 632-639.

**EFFROS B.** 2003  
*Merovingian Mortuary Archaeology and the Making of the Early Middle Ages*. University of California Press, 296p.

**EFFROS B.** 2012  
*Uncovering the Germanic Past: Merovingian Archaeology in France, 1830-1914*. Oxford University Press, 427p.

**GODOY-LORITE A., GUIMERÀ R., MOORE C., SALES-PARDO M.** 2016  
*Accurate and scalable social recommendation using mixed-membership stochastic block models*. PNAS, 113(50), p. 14207–14212.

**HÄRKE H.** 1990  
« *Warrior Graves* » ? *The Background of the Anglo-Saxon Weapon Burial Rite*, Past & Present, 126, p. 22-43.

**KNÜSEL C., RIPLEY K.** 2000  
« The Berdache or man-woman in Anglo-saxon England and Early Medieval Europe » dans FRAZER W.O., TYRRELL A. (éd.) - *Social identity in Early Medieval Britain*, Londres, New York, 2000, 157-192.

**LUCY S.** 1997  
« Housewives, warriors and slaves ? Sex and gender in Anglo-Saxon burials » dans MOORE J., SCOTT E. (éd.), *Invisible people and processes : writing gender and childhood into European archaeology*, Londres, New-York, 1997, 150-168.

**PERIN P., BUCHET L.** 2023  
« Quelques réflexions sur les tombes mérovingiennes et l'archéologie du genre » dans NALLBANI E. (éd.) - *Hommage à Michel Kazanski*, Paris, Beograd, 131-149 (Travaux et Mémoires, 27).

MURAIL P., BRUZEK J., HOUET F., CUNHA E. 2005  
DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements, *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, 17 (3-4), pages 167-176.

POUX-MÉDARD G., VELCIN J., LOUDCHER S. 2021  
*Information interactions in outcome prediction : Quantification and interpretation using stochastic block models*. Fifteenth ACM Conference on Recommender Systems (RecSys), p. 199–208.

POUX-MÉDARD G., VELCIN J., LOUDCHER S. 2022  
*Serialized interacting mixed membership stochastic block model*. IEEE International Conference on Data Mining (ICDM), p. 1143-1148.

SEBILLOTTE-CUCHET V. 2015  
*Avant-Propos, Les nouvelles de l'archéologie*, 140, 2015, 3-4.

#### Pour citer cet article

POUX-MEDARD G., ABERT F., BARRAND EMAM H., CHENAL F., FISCHBACH T., SCHNEIDER E., TOUZET A. 2025

« Identification de l'âge et du sexe des défunt à partir du mobilier funéraire mérovingien : une approche d'apprentissage automatique » dans LE FORESTIER (dir.), *Les ensembles funéraires altomédiévaux : nouvelles recherches et nouvelles méthodes*, actes du colloque des 42<sup>e</sup> journées internationales de l'AFAM. Saint-Germain-en-Laye (Yvelines), du 5 au 8 octobre 2022, Sevran, Association Française d'archéologie mérovingienne et Association Archéologie des nécropoles, p. 385 à 404.



Les 42<sup>es</sup> journées internationales d'archéologie mérovingienne (Saint-Germain-en-Laye, 5-8 octobre 2022) ont permis de présenter les nouvelles recherches engagées depuis près de deux décennies sur les ensembles funéraires altomédiévaux, grâce en particulier au renouvellement des méthodes. Elles ont ainsi donné l'occasion de faire un bilan des trois projets collectifs de recherche menés parallèlement en Île-de-France, en Champagne-Ardenne et en Alsace, qui ont eu notamment pour objectif de faire un état des connaissances, notables, acquises dans chacune des régions par le développement de l'archéologie préventive. D'autres recherches plus locales ou sur des sujets plus spécifiques ont également été exposées, permettant d'aborder diverses thématiques comme l'architecture des tombes, le recrutement des défunt, les lésions osseuses ou le mobilier funéraire.

Ces journées ont enfin été l'occasion de montrer les perspectives nouvelles ouvertes par les archéosciences, par le biais de disciplines et de méthodes aussi variées que la paléogénétique, l'étude des restes organiques, l'enregistrement numérique exclusif ou encore le *machine learning*.

Cette publication rend ainsi compte de la diversité des approches adoptées aujourd'hui dans l'étude des ensembles funéraires altomédiévaux, en mettant en évidence leurs acquis ainsi que, pour certaines, leurs limites.



Inrap<sup>+</sup>

Institut national  
de recherches  
archéologiques  
préventives



ISBN : 979-10-415-6278-7

Prix 30 €



979-10-415-6278-7