Chapitre 2.1.2
Faire parler les artefacts et les vestiges mobiliers pour les recherches de provenance
Formalisation des données

Stéphanie LEROY(1), Marie ORANGE(2), Ludovic BELLOT-GURLET(3), François-Xavier LE BOURDONNEC(4)

(1) IRAMAT UMR 7065, CNRS/NIMBE UMR 3685, CEA, CNRS, Université Paris-Saclay, Gif-sur-Yvette (2) Archéorient UMR 5133, CNRS, Maison de l'Orient de la Méditerranée, Lyon/Archéosciences Bordeaux UMR 6034, CNRS, Université Bordeaux Montaigne, Université de Bordeaux, Pessac (3) de la Molécule aux Nano-objets : Réactivité, Interactions et Spectroscopies, MONARIS UMR 8233, Sorbonne Université, CNRS, Paris, France (4) Archéosciences Bordeaux UMR 6034, CNRS, Université Bordeaux Montaigne, Université de Bordeaux, EPHE-PSL, Pessac

Fragment de roche, tesson céramique, artefact métallique… autant de vestiges mobiliers que l'on retrouve ici et là dans le paysage archéologique en fonction des différents types de mobilités induits par les comportements humains du passé. Les études en sciences archéologiques s'y intéressent et celles portées sur la *provenance* en particulier cherchent à retrouver trace de leur origine, celle de leur genèse et de leur histoire matérielle. Ces études de provenance peuvent être destinées à identifier la nature des matériaux sources* ainsi que les lieux* et milieux (gîte minéral, couche d'argile, etc.) qui ont été exploités pour fabriquer les vestiges. Elles peuvent également chercher à identifier les ateliers où il y a eu transformation (métamorphose ou modification de forme) des matériaux primaires utilisés pour les créer. Dans d'autres cas, il peut également s'agir de retrouver les lieux par lesquels les artefacts ont transité avant d'être utilisés. Ce sont alors des indices sur la circulation d'objets ou de populations qui sont obtenus, et à travers ceux-ci, ce sont des lieux géographiques, identifiés comme lieux de genèse, de transit, d'usage ou d'abandon, qui peuvent être mis en relation. Lorsque ce type d'observation est réalisée dans la diachronie, des changements de relations peuvent être saisis dans le temps et la dynamique des interactions – celles témoignant de réseaux de diffusion ou de mobilités – peut être approchée.

Retrouver trace de l'origine des vestiges est possible car les propriétés intrinsèques des matériaux sources et ressources qui ont été utilisés sont conservées ou transmises d'une manière ou d'une autre entre les différents états transformés de la matière d'un même ensemble de production matérielle (*cf.* 2.1.1), des objets ressources aux objets mobiliers. Cette transmission induit un effet de lien, une sorte de continuité chimique identifiable entre les vestiges qu'il est possible d'exploiter pour mettre en correspondance *a posteriori* des éléments qui se sont retrouvés déconnectés. Une source de propriétés largement mobilisée est la composition chimique en ce qu'elle permet d'accéder à une multiplicité de marqueurs identitaires (sous-entendu signant) du vestige lui-même. Il est aussi envisageable, si l'on monte en généralité, de cerner la signature d'un environnement* en considérant un ensemble de vestiges. Dans les études de provenance, ce sont surtout à ces ensembles plutôt qu'à des individus isolés que l'on s'intéresse, et face à la question qui est posée, la signature* peut se définir à différents niveaux d'agrégation des vestiges. Ainsi mobilisée en tant que schème ontologique, la signature trouve plusieurs dénominations dans les exemples qui suivent, chacune faisant appel à un concept propre de mise en relation : « signature (géo)chimique », « empreinte digitale »/« fingerprint », « signature d'une production », « signature d'un atelier », « signature d'un gisement », etc.

Vestiges, ensembles, assemblages* sont des éléments de contexte qui opèrent tous dans les formes ou les géométries finales de ces signatures. Les outils de l'analyse statistique permettent de les approcher et d'en étudier les correspondances vis-à-vis des liens dont elles témoignent. Il est pour cela nécessaire de construire une grille de lecture qui fasse dialoguer information historique et représentation des données. L'enjeu consiste alors à utiliser des données observables[1] en capacité d'incarner les signatures et, comme le décrit l'auteure du Chapitre Introductif, à « traduire les questions archéologiques en questions statistiques » afin d'utiliser les outils en correspondance avec les objectifs fixés.

1 Les composantes de l'observation statistique

Les composantes qui interviennent dans l'analyse statistique pour les études de provenance sont les suivantes. Il y a deux catégories d'observables : (1) les « individus matériels* » et (2) les « marqueurs chimiques » qui sont sources d'informations historique et chimique des signatures d'intérêt. Il y a ensuite la catégorie (3) des « populations* statistiques » qui, soumises à l'analyse, sont des objets de comparaison élaborés à partir des observables précédents et propres à documenter une source en particulier.

Dans la catégorie des observables, les individus matériels jouent un rôle prioritaire. Il peut s'agir de vestiges géologiques ou archéologiques, qu'ils soient mobiliers (un fragment de roche taillée, un tesson céramique, un artefact, etc.) ou non (comme une roche primaire) qui sont collectés sur le terrain. Des éléments de vestiges peuvent aussi être considérés comme individus (par exemple des inclusions d'intérêt au sein des métaux ferreux *cf.* 2.1.4). Ces observables sont sélectionnés avant tout vis-à-vis de leur capacité sémiotique à représenter un **lieu** à différentes échelles possibles : une lo-

[1] Nous reprenons le terme d'observables tel qu'il est défini dans le Chapitre Introductif de l'ouvrage (Cf. Chapitre Introductif. Dialogues entre concepts archéologiques, objets empiriques et méthodes statistiques).

calisation intra-site (couche stratigraphique, unité archéologique, etc.), un site (gîte, atelier, édifice, endroit d'usage), un ensemble de sites voisins, ou une zone géographique encore plus étendue jusqu'à pouvoir représenter une région entière si l'étude le demande. La notion de lieu intra-objet avec l'exemple des inclusions peut également être considérée. Censés définir plus largement un environnement spatial, les points de collecte des vestiges sont donc loin de représenter uniquement de simples coordonnées géographiques (X,Y,Z). Il est donc essentiel de comprendre le lien que l'individu entretient avec son milieu de découverte, ce que permet une étude de l'objet et du terrain orientée vers les enjeux de taphonomie des sites, de géologie et/ou de géomorphologie des milieux.

Le deuxième type d'observables relève des propriétés chimiques utilisées pour donner 'corps' aux signatures cibles et qui constitueront les variables de l'analyse statistique. Ce sont des marqueurs de composition chimique (élémentaires ou isotopiques) lus dans les matériaux des individus et comparables, d'un état à un autre, entre les vestiges. Le choix des propriétés chimiques considérées dépendra directement du type de production matérielle (céramiques, métaux, obsidiennes, etc.) et de la question qui est posée. Pour le cas de l'obsidienne et des matières colorantes, où l'enjeu de la comparaison cible systématiquement des géomatières[2] qui n'ont pas subi de transformation chimique anthropique, seront mobilisées les valeurs absolues de concentrations élémentaires. Pour les métaux cuivreux, ce sera un rapport de concentrations isotopiques permettant de faire le lien entre minerais et objet métallique. Pour les métaux ferreux, ce peut être des valeurs absolues de concentrations élémentaires si des minerais sont comparés entre eux, ou encore des rapports de concentrations élémentaires si la comparaison se fait entre minerais, scories et inclusions dans l'objet, c'est-à-dire ici des matériaux de nature différente.

Les objets de comparaison qui sont mis en relation dans l'analyse statistique sont les signatures de phénomènes qui sont assignables à la fois aux lieux et à l'activité humaine : l'exploitation d'une ou plusieurs ressources, une manière de produire sur un atelier, l'utilisation de matières premières sur un atelier, etc. Construites à partir d'un ensemble de marqueurs chimiques identitaires, plusieurs dénominations servent à y faire référence dans les exemples qui suivent : « signature d'un artefact », « signature d'un ensemble d'objets », « signature d'un atelier de production », « signature d'un gîte minéralogique », « signature d'une pratique », etc. Ces objets de comparaison sont élaborés à partir d'un assemblage d'individus*, et qui, par regroupement, définissent une population statistique. Celle-ci peut être représentée par un même type de vestiges, comme dans le cas de l'obsidienne et des matières colorantes où seules des géomatières sont étudiées, l'objectif étant de rechercher l'origine géologique de fragments distancés de lieux-sources primaires. Il s'agit également d'un même type de vestige lorsque l'on considère la signature d'un ensemble d'objets mobiliers que l'on cherche à relier à une ressource. La population statistique peut aussi être construite à partir de vestiges aux différents états de transformation de la matière, comme pour l'exemple des métaux, où minerais, déchets de production et objets métalliques peuvent servir à traduire la signature d'une production associée aux ateliers. Dans ce cas, il sera important de

[2]Nous reprenons le terme de géomatières tel que mentionné dans l'exemple des matières colorantes, *cf.* 2.1.4.

construire la signature à partir de marqueurs chimiques communs à l'ensemble considéré pour rendre possible l'assemblage ciblé. Au final, les populations ainsi construites affectent les signatures à la fois par de la distribution physique (au sens matériel) et de la variabilité chimique, sans que ces deux facteurs ne soient nécessairement corrélés. Autrement dit, ce n'est pas parce qu'une population est représentée par un nombre plus grand d'individus que sa représentation chimique (et donc sa dispersion) sera plus étendue (au sens où il y aurait « somme » des variabilités). La représentativité d'une signature ne se joue pas uniquement à l'échelle du nombre d'individus la nourrissant mais aussi de la capacité de représenter la spécificité chimique de l'environnement. Ici, nous sommes face à une contrainte limitante pour les études de provenance : la complexité des variations chimiques présentes dans l'environnement ne peut être connu *a priori* malgré l'analyse la plus précise possible du terrain. Si l'on considère l'environnement du gîte minéral par exemple, des variabilités chimiques peuvent être fortement prononcées sur des distances très faibles. À l'inverse, il est possible que des sources primaires, pourtant éloignées de plusieurs dizaines ou centaines de kilomètres, soient caractérisées par des compositions chimiques très proches. Cette observation peut être également projetée sur les signatures « production » au sein d'ateliers qui, puisqu'elles dépendent de traditions techniques et donc de modes opératoires, ne sont pas nécessairement « standardisées » au point d'aboutir à une composition d'ensemble homogène. Pourtant, ces signatures sont utilisées comme un « tout » représentant un **environnement** spécifique (associé aux contextes spatial – composante géologique incluse – et/ou humain). Il est donc important de saisir l'importance du rapport entretenu entre « origine source » et « unité spatiale », que nous nommerons unité* source, qui correspond en quelque sorte au groupement chimique associé à un environnement. Ainsi, une population statistique n'est pas nécessairement une unité source, et c'est en tout cas une vérification à laquelle il faut pouvoir procéder à travers la lecture des échelles de variabilité des signatures. La forme multidimensionnelle des données analysées est souvent appréhendée une fois l'analyse multivariée effectuée et à l'aide de celle-ci.

Et le « Quand » ? En l'état, la méthode de comparaison de données n'inclut pas de critères temporels. La possibilité d'analyser les relations entre des signatures pour des faits du passé qui ne seraient pas contemporains n'est donc pas exclue. Pour certains exemples, les échelles de temps intégrées dans les profils des signatures (un profil géologique par exemple) peuvent être suffisamment étendues pour ne pas avoir à se poser la question (concept source propre à l'exemple 2.1.3). Pour d'autres, les individus peuvent être sélectionnés au regard des dates qui leur sont attribuées, permettant par la même occasion de venir contraindre, au préalable à l'analyse, les interprétations de mise en relation (*cf.* 2.1.7). Mais dans la majeure partie des exemples, ce sera la contextualisation dans le temps des relations spatialisées qui permettra d'éprouver le raisonnement final. En faisant abstraction du temps dans la méthode statistique, le résultat final peut aussi amener à réfléchir sur la capacité d'un observable donné à représenter la signature propre au phénomène passé, qui lui a nécessité d'être circonscrit dans le temps et dans l'espace.

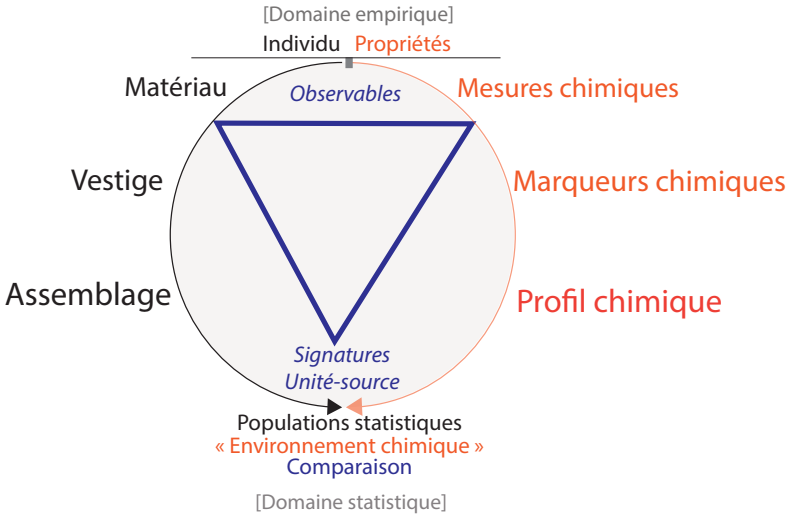


FIG. 1 : Les catégories de composantes qui interviennent dans les analyses statistiques pour les études de provenance.

2 Ajuster sa focale d'observation pour élaborer les signatures

Le point de départ de toute analyse statistique est une problématique de nature historique qui intègre le plus souvent des échelles (spatiales et/ou temporelles) étendues sur les faits du passé, ainsi qu'une diversité, si ce n'est une complexité, des environnements chimiques de lieux d'intérêt. Mobiliser l'ensemble des données dans le cadre d'une seule et même analyse est généralement risqué. En effet, les informations chimiques présentent des dimensions relationnelles multiples (propres aux caractéristiques environnementales et humaines) dont la combinaison crée un champ de variabilités à la structure finale souvent indiscernable. Dans l'analyse statistique, ce sont des variabilités aux ordres de grandeur comparables qui permettent d'étudier les correspondances entre les signatures. Si les ordres de grandeur des signatures n'expriment pas les mêmes sources de variabilité à l'origine, la comparaison est possible mais la relation sera moins directe, et un effort de retour sur données initiales sera indispensable. Pour le cas des obsidiennes (*cf.* 2.1.3), l'environnement des signatures sera toujours le même, l'objectif étant d'étudier les correspondances entre les matériaux primaires et les fragments distancés des gîtes primaires. Quelle que soit la question posée, les vestiges sont souvent de même nature ; les apports exogènes ou pertes chimiques sont rares, tous reflètent donc le même processus (voire le même phénomène) à l'origine. Ainsi, la comparaison des signatures opère dans un rapport direct, raison pour laquelle on peut parler dans ce contexte particulier de concept « source ». La problématique est toute autre dès lors que des chaînes de transformation des matériaux sont mises à l'œuvre pour produire les matières utilisées, ce qui est le cas pour les matières colorantes (*cf.* 2.1.4), les métaux (*cf.* 2.1.5 et 2.1.7) et les céramiques (*cf.* 2.1.6). Pour ces exemples, des dimensions de nouvelles variabilités assignables à des combinaisons d'information chimique (apports exogènes propres à

la technique, mélange, recyclage), ou perte endogène, à des effets de dilution, ou à des effets de taphonomie (transformations naturelles *a posteriori*), peuvent être intégrés dans le schéma statistique des mises en relation. On peut penser que le niveau d'information sera d'autant plus mélangé que le nombre de faits impliqués dans la production matérielle[3] sera important. Ainsi, on peut tenter d'être exhaustif sur la signature d'un atelier, auquel cas sera considéré un grand nombre et de vestiges et de valeurs chimiques en attribution pour témoigner de l'activité mise en place sur ce lieu. Il ne faudra cependant pas oublier que la signature finale obtenue sera l'empreinte de l'ensemble des contributions matérielles, des sources géologiques commes des ressources techniques, selon les procédés mis en œuvre pour la production, et témoignera donc aussi largement de pratiques plutôt qu'uniquement de ressources en particulier. Par exemple, les signatures utilisées pour les études de provenance sur les tessons céramiques sont le résultat de contributions issues de ressources argileuses, de transformations et d'ajouts. Ils présentent donc une composition chimique résultant d'une combinaison de facteurs, témoignant moins d'une ressource unique que d'une tradition de fabrication (c'est-à-dire incluant pratiques techniques et gestion des ressources en matières premières). Pour les métaux, fer ou cuivre, il est possible de procéder à la sélection d'éléments chimiques propres à pouvoir documenter la provenance à un atelier de production ou à des ressources minérales. Dans la comparaison, les variations sont observables directement que si les signatures sont du même type. Face à l'exhaustivité, surtout chimique, mentionnons également qu'une attention particulière doit être portée sur la capacité discriminante qui se doit d'être conservée par les signatures. L'exhaustivité chimique sera toujours favorable à une variabilité parfois « non significative » et donc limitante pour l'observation des relations entre les signatures.

L'ajustement de la focale d'observation se joue également à une autre échelle, beaucoup plus fine : celle des hétérogénéités qui peuvent caractériser le matériau lui-même. C'est un profil chimique propre à des composants du matériau qui est recherché et qu'il faut donc pouvoir cibler dans l'acquisition des observables. La question est récurrente pour des vestiges composites comme les artefacts ferreux, les matériaux céramiques (constitués de minéraux argileux certes, mais aussi d'autres composants variés combinés et de forme inclusionnaire) ou encore celui des matières colorantes archéologiques (dont les auteurs de la partie 2.1.4 mentionnent la nécessité de comprendre les caractéristiques pétrologiques avant toute analyse statistique).

Ainsi, c'est une échelle de construction qu'il faut pouvoir mettre en place, abordée dans un rapport juste entre restitution historique et représentation des données, avant le traitement de ces dernières avec les outils de la statistique. Il est donc souvent nécessaire de décomposer son problème afin de délimiter le champ des questions pour cibler avec précision le phénomène passé sur l'origine que l'on cherche à atteindre et donc les processus chimiques qui en témoignent et que l'on cherche à mettre en relation. Est-ce que je m'intéresse en priorité à la circulation des objets ? À la diversité et/ou la particularité des ressources exploitées ? À la mise en évidence d'un contact en particulier, comme celui qui pourrait exister entre une zone de production et une

[3] Nous faisons le choix de ne pas parler de « chaîne opératoire », car nous nous attachons à analyser uniquement les relations matérielles sans aussi considérer les étapes d'un processus de fabrication d'un objet.

destination cible des productions ? Ces questions ne sont que quelques exemples des interrogations qui peuvent se poser dans les études de provenance. Chaque question propre à la détermination d'origine fait référence à une particularité de lieu(x) et d'activités humaines à la fois. Poser la question aidera à ajuster la focale d'observation, à choisir les observables représentatifs et à engager l'observation des relations dans un rapport le plus direct possible. En pratique, la construction de la grille de lecture sera contrainte par les possibilités de collecte des artefacts étudiés, vis-à-vis de leur reconnaissance (présence/absence), d'une accessibilité aux sites, voire par l'état des connaissances que porte l'observateur sur les contextes archéologiques et géologiques. Même si la sélection se veut être la plus exhaustive possible, le résultat final sera souvent lacunaire, et la capacité des observables à représenter la signature des environnements d'intérêt toujours imparfaite. Ainsi, les interprétations proposées devront toujours être pesées en fonction des apports mais aussi des limites que l'observateur sera en mesure de soulever. Une étude inscrite dans la durée sera favorable à la reconnaissance des ensembles formés en alimentant les jeux de données pour enrichir les interprétations et « mettre à l'épreuve » les définitions des ensembles.

3 Le passage de l'observable à l'unité source

L'enjeu consiste aussi à transformer les observables en matrice(s) de données qui représenteront les signatures chimiques et/ou les unités sources, et le passage de l'un à l'autre n'est pas direct. Les données primaires à ces matrices sont les concentrations des éléments chimiques mesurées dans le matériau et en capacité de témoigner du lien ciblé. Lorsque le lien de comparaison cible la composition chimique, les concentrations chimiques seront utilisées directement, dans d'autres cas, la mise en correspondance concerne des matériaux dont les compositions ne sont pas directement comparables, à l'image des exemples sur les métaux (*cf.* 2.1.5 et 2.1.7), pour lesquels seront utilisés des rapports de corrélation de valeurs élémentaires. Dans tous les cas, les données initiales, continues, sont acquises à l'aide de techniques d'analyses chimiques qui seront globales ou ponctuelles (MEB-EDS, ICP-MS, LA-ICP-MS, XRF, INAA, etc.), en fonction du matériau et de la question qui est posée. Par exemple, le LA-ICP-MS sera utilisé pour analyser des inclusions d'ordre millimétrique au sein d'un objet (telles que des inclusions d'obsidienne utilisées comme dégraissant dans certaines céramiques, ou des inclusions de scorie dans les objets ferreux), alors que des techniques d'analyse plus globales comme l'XRF seront choisies pour l'étude de tessons céramiques ou d'artefacts en obsidienne. Si l'usage de techniques analytiques est pluriel, une intercomparabilité des données acquises devra être éprouvée par ailleurs.

Les mesures obtenues sont ainsi exprimées dans des ordres de grandeur qui correspondent à des échelles différentes (généralement exprimées à la fois en pourcents (%) et en parties par million (ppm – 0,0001%)). Lorsqu'il est fait appel à la composition totale et que des valeurs absolues de teneurs élémentaires sont utilisées, les deux catégories de grandeur (% et ppm) sont regroupées, et les valeurs les plus grandes auront alors un poids supérieur dans la restitution des variabilités observées. C'est-à-dire que les variabilités témoignées par les valeurs les plus faibles seront plus « ramassées » par rapport aux variabilités des valeurs les plus grandes. Ramener les ordres de grandeur à l'échelle de l'unité est donc essentiel pour assurer un même ordre de comparaison,

ce que les opérations de normalisation* des données rendent possible (*cf.* 4). Dans le cas où ce sont des rapports élémentaires qui sont ciblés, une transformation des données sur les valeurs de composition peut être appliquée : du simple outil d'une construction de corrélation deux à deux, à une transformation plus généralisante dès lors que la dimension multivariée est prise en compte (*cf.* 6). Cette transformation permet – plus indirectement – d'accéder à l'enjeu de la normalisation des données, puisque les rapports de valeurs sont exprimés dans un même ordre de grandeur. Il n'est cependant pas exclu que les différents rapports de corrélation aboutissent à des variabilités si différentes qu'une étape de normalisation sur ces valeurs s'avère aussi nécessaire. Enfin, notons que pour des raisons de limites de détection ou d'absence d'acquisition, certaines mesures élémentaires peuvent être manquantes au sein des populations statistiques. Dans le domaine des mesures compositionnelles, puisque les données sont continues, il n'est cependant pas possible d'intégrer des identités au profil chimique incomplet, c'est-à-dire aux variables non renseignées. Des solutions de remplacement existent (*cf.* 4), mais il est important de savoir que ces dernières entraînent nécessairement des écarts possibles au profil originel, modifiant ainsi les formes des signatures. Il est toujours envisageable de supprimer les individus qui sont concernés, mais ce geste peut alors impliquer la suppression d'informations susceptibles de documenter une partie de l'identité chimique. Ainsi, conserver ou supprimer les individus exprimés par de l'information lacunaire n'est pas un choix évident, et l'observateur ayant éprouvé le résultat des deux tests sera sans doute en capacité de proposer la « meilleure » solution. De manière générale, une étape de « nettoyage des données » se révèle donc souvent indispensable alors qu'elle est, notons-le, très peu souvent mise en pratique (ou de manière subsidiaire). Un retour à la fois sur la collecte et sur l'acquisition des données analytiques est souvent une aide non négligeable à la décision.

Les signatures ainsi alimentées par les populations statistiques, une étude descriptive de leur état de granularité* se révèle nécessaire afin de lire la cohésion des données pour chacune d'entre elles. Cette observation peut être réalisée à l'aide des indicateurs statistiques des représentations univariées (par exemple, moyenne, médiane, écarts-types, distributions, *cf.* 4), de la description par l'analyse bivariée (*cf.* 5), voire en utilisant directement les analyses factorielles (*cf.* 6), selon le nombre de marqueurs chimiques pris en considération et la capacité de l'observateur à lire les spécificités techniques pour chaque représentation statistique. En observant les variabilités au sein d'une ou plusieurs populations statistiques, il est possible de préciser les champs de délimitation de chaque signature et vérifier qu'elle puisse être associée à une unité source, entendue ici comme circonscrite à la fois géographiquement et géologiquement. La représentation d'une unité source dépend donc plutôt de sa capacité à être distincte des autres que d'une quantité de données disponible/traitée. Les exemples présentés dans les chapitres qui vont suivre permettront d'illustrer différentes situations. Dans tous les cas, il s'agit de faire jouer des ensembles d'individus matériels et de variables avec un juste équilibre entre les conditions requises sur le plan quantitatif, pour limiter les biais et les contraintes thématiques nécessaires pour que l'analyse garde du sens eu égard à la question posée.

L'analyse descriptive de chaque ensemble aidera également à identifier les observables qui se détachent des tendances plus largement observées, que nous nommerons « épiphénomènes ». Ces situations ne sont pas rares. Sur le terrain, il est possible de retrouver des vestiges décontextualisés de leur position primaire, car déplacés *a posteriori* (par des procédés naturels ou des transports), parfois à des distances très lointaines. Il est même possible de les retrouver abandonnés et mélangés à d'autres vestiges de nature similaire sur des lieux qui se trouvent être également un centre d'intérêt pour l'étude de provenance. L'ensemble des exemples abordés dans les chapitres qui suivent font état, d'une manière ou d'une autre, de cet écueil. Si la connaissance du terrain permet de ne pas sélectionner les cas les plus évidents dès le début de l'étude, il n'est pas toujours simple de les éviter, et certains peuvent se retrouver soumis à l'analyse statistique. Dans le meilleur des cas, il sera possible de les repérer grâce à des caractéristiques chimiques qui se distinguent de l'ensemble, d'une spécificité attendue, et il sera alors possible de les exclure. Dans d'autres cas, cette donnée peut se retrouver mélangée à celles représentant d'autres faits, et à cet égard, induira des écarts possibles à la représentation dont la forme finale en sera modifiée ainsi que l'interprétation des résultats.

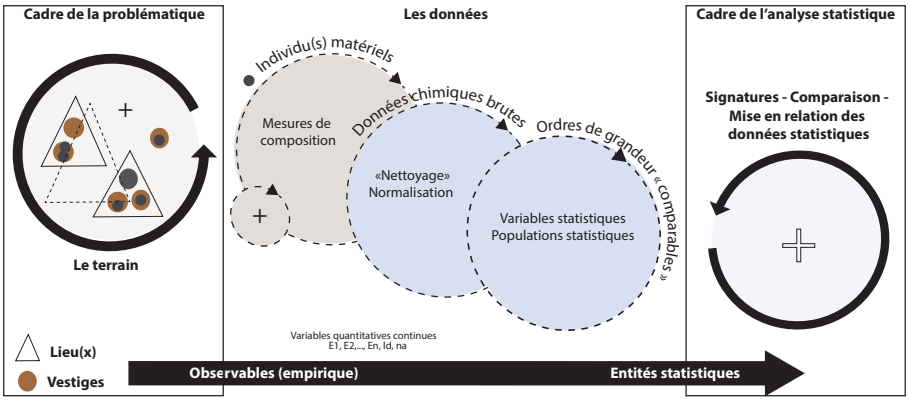


FIG. 2 : Étapes d'analyses entre les observables et la mise en relation statistique des processus.

4 Comparer les signatures pour étudier les correspondances

4.1 Quelles variables et quelles données ?

Ce sont alors différents ensembles chimiques qui sont obtenus et dont il faut pouvoir étudier les correspondances en lien avec les relations qu'ils pourraient entretenir. Le problème est ici pragmatique, il s'agit d'une comparaison d'ensembles-signatures rendue possible en utilisant les outils de la statistique. Si un travail descriptif sur les signatures a été réalisé en amont, impliquant un choix de variables et de données permettant de les définir, dès lors que l'on s'intéresse à une nouvelle comparaison, l'observation opère dans un nouveau référentiel qui permet une mise en interaction nouvelle des éléments de l'analyse. Dans certains cas, il sera possible de se poser à nouveau la question de la capacité de l'observable donnée à représenter la signature : les

données s'écartent des tendances observées et des épiphénomènes qui avaient échappé à l'observation jusqu'alors sont repérés. Le choix de les retirer ou de les conserver n'est pas toujours simple et la décision finale revient souvent à l'expérience de l'observateur qui connaît le terrain et qui aura pris le temps de « jouer » avec les données, de tester différentes mises en relation, avec ou sans certaines variables ou individus. En pratique, une observation par va-et-vient dans la comparaison sera une approche efficace pour une lecture affinée – et *a posteriori* - des ensembles. Dans le meilleur des cas, une redistribution bénéfique des individus au sein des différents ensembles pourra opérer lorsque le contexte historique permet éventuellement de le penser. Certaines données peuvent même devenir une unité source à part entière (les artefacts deviennent nouvelle population). C'est donc parfois au-delà du terrain, et à travers un travail d'interprétation en lien avec le contexte historique, que l'individu peut être relié au concept (*cf.* Chapitre Introductif). Dans d'autres cas, et selon les données à disposition, il est également envisageable de conserver une source possible d'épiphénomène dans l'observation (*cf.* 2.1.7, exemple de la Figure 3). On se laisse alors la possibilité de revenir sur son interprétation en fonction de l'évolution de l'étude.

Le choix des variables relève avant tout de leur apport sur le plan thématique, c'est-à-dire vis-à-vis de leur capacité à décrire les processus que l'on cherche à mettre en relation (témoignent-elles du même type d'origine – géologiques et/ou humaines ?). Il ne faut cependant pas oublier que leur rôle est également d'aider à discriminer les populations d'origine différente. Prendre davantage de variables descriptives peut parfois augmenter la variabilité au point de réduire la capacité à différencier les ensembles. Aussi, il faudra bien saisir l'apport de chacune des variables à la représentation de la connexion qui dépend elle-même d'un lieu. Ici, les enjeux de représentation de lieux et des connexités à suivre peuvent être complexes. Parfois, il est donc plus intéressant de sélectionner des variables que l'on sait être discriminantes pour des zones précises, et ensuite vérifier à nouveau – et au cas par cas – que l'ensemble descriptif avec plus de variables intégrées peut correspondre. L'analyse peut donc souvent se faire en plusieurs étapes. Il ne faut pas perdre de vue que la pratique joue un rôle important dans la mise en œuvre de ce processus, car elle permet de mettre en interaction l'ensemble des choix qui sont opérés et, par les changements qu'ils induisent, de mieux cerner leur rôle dans l'analyse. Enfin, il est aussi possible de revenir en itération sur les données primaires, pour une interrogation sur la validité des mesures ou la pertinence de la sélection de certaines variables. Cette « compréhension » des données issues de leur manipulation participe d'une façon générale à l'interprétation du sens ou de la pertinence des données analytiques avec un retour sur les problématiques abordées ou les étapes de transformation des matériaux considérés.

Dans le cadre de la comparaison, un critère chimique sélectionné en tant que variable pour l'analyse statistique implique que l'ensemble des échantillons soumis à l'analyse le possède, ce qui en théorie est le cas. Dans la pratique, l'observateur fait souvent face à des limites contraignantes à ce sujet, en lien avec l'évolution et le choix des techniques et des méthodes d'analyse. Les études de provenance, d'autant plus lorsque ces dernières s'inscrivent dans le temps long, nécessitent de faire appel à des jeux de données dont certains d'entre eux seront marqués par des lacunes en information, ce

qui est d'autant plus probable lorsque des résultats d'analyses plus anciennes sont réinvestis pour l'étude en cours.

4.2 Mettre en correspondance pour assigner une provenance

Dans les études de provenance, il existe un caractère assignable. Ainsi, une identité chimique inconnue est généralement comparée à ce qui est connu, c'est-à-dire à une référence qui sert avant tout à nommer l'origine (détermination d'origine). Ce type de population statistique est généralement nommé **Référentiel** dans les études de cas qui suivent. En tant que population statistique propre à pouvoir refléter l'empreinte d'une ou de plusieurs unités sources, le référentiel est le résultat d'une utilisation choisie et réflexive d'un échantillonnage et de données représentatives. Dans le contexte spécifique de l'analyse statistique, il ne faut ainsi pas confondre « base de données » et « référentiel ». Si les bases de données sont des architectures « froides », « figées » dans leur structure et utilisées pour construire des populations statistiques, les référentiels sont de nature « plastique » qui peuvent évoluer en fonction des problématiques abordées et de l'évolution de la recherche puisqu'ils sont populations statistiques. Un référentiel n'est pas un support indispensable à la comparaison. En effet, on peut trouver intérêt à comparer uniquement des ensembles de vestiges inconnus afin de déterminer si leurs identités chimiques se différencient ou non entre elles et donc rechercher s'ils appartiennent au même ensemble ou pas. La relation avec l'origine est plus indirecte mais des indices sur les spécificités de provenance sont malgré tout obtenus.

C'est la comparaison de chaque empreinte issue des signatures et l'observation des ressemblances suggérées par l'analyse qui prévaut dans les méthodes qui ont trouvé leur place dans la détermination d'origine. Plusieurs approches sont possibles. Les analyses factorielles (*cf.* 6, essentiellement l'Analyse en Composantes Principales – ACP) peuvent être utilisées afin de décrire la structuration des données dans leurs ensembles. Ces analyses seront également utiles pour observer des particularités spécifiques aux variables retenues (corrélations possibles ou pouvoir discriminant de certaines). L'observateur sera alors amené à visualiser le positionnement des individus dans l'espace multivarié et à interpréter des proximités possibles entre des données (*cf.* 2.1.3 à 2.1.7). Plus indirectement, il s'agit d'évaluer les rapprochements entre des signatures dont les domaines d'existence seront délimités par l'observateur lui-même, fort de la connaissance qu'il a du terrain et de la représentativité qu'il associe aux données. Il faudra donc être attentif aux résultats obtenus et aux contributions respectives des variables dans différents plans factoriels, chacun ne « représentant qu'une partie de la réalité » (*cf.* 6). D'autres analyses qui font appel à des concepts de classification (par exemple la Classification Ascendante Hiérarchique – CAH, *cf.* 9) amènent également des solutions pour comparer les signatures, en proposant d'évaluer les discontinuités entre les données (*cf.* 2.1.4 à 2.1.6). Ce type de construction est alors bâti sur un ensemble de critères qui seront jugés pertinents par l'observateur et sur des enjeux de cohésion interne des données et de leur isolement possible pour la mise en évidence de différents ensembles (*cf.* 9). Cette étape suppose néanmoins d'être en capacité de reconnaître les contours des phénomènes soumis à l'analyse, c'est-à-dire ceux qui procèdent d'une échelle de construction supervisée dans l'assemblage des in-

dividus matériels. Cette appréhension des données peut aussi être favorable à l'usage d'outils de type classification supervisée à l'image de l'Analyse Discriminante à visée prédictive (*cf.* 10) ou selon une approche strictement géométrique (*cf.* 10, §3.2) qui permet de prendre en considération le nombre de signatures ainsi que les métriques associées aux variabilités intra-signature (*cf.* 2.1.4 et 2.1.5). L'enjeu de la comparaison cible alors les variables explicatives qui permettent de discerner le plus possible ces signatures. Au-delà d'une démarche de lecture graphique des données, on peut aussi chercher à fournir une approche chiffrée des correspondances entre individus dont on cherche à identifier l'origine et les classes du référentiel en comparaison (exemple de la méthode des « k plus proches voisins », *cf.* 10 et 2.1.7). La pertinence de l'application des approches à visée prédictive dépendra alors de l'état des connaissances sur les possibilités sources en entrée de l'analyse et aussi de l'échelle d'étude géographique investie dans l'étude. N'oublions pas que la mise en relation statistique opère dans un référentiel propre aux observables soumis. Autrement dit, ce qui n'est pas représenté est absent de la logique statistique, laissant l'opportunité au calcul mathématique de proposer « avec assurance » des attributions par défaut. Pourtant, l'idée que des ensembles de données sources ne soient pas inclus dans l'analyse (une source épuisée, inaccessible de nos jours ou encore non repérée dans le paysage archéologique) est tout à fait envisageable, et même, probable dans de nombreux exemples. Cet état partiel des connaissances est le fruit récurrent de biais méthodologiques dans les attributions de provenance. S'il est possible de « prédire » (au sens chiffré) une appartenance à ce qui est représenté, il n'est pas possible de le faire pour ce qui n'existe pas. C'est également vis-à-vis de ce concept de l'inconnu que les conclusions restent de manière générale temporaires, laissant la possibilité qu'une nouvelle source qui correspondrait également ou même davantage à la signature de l'artefact soit ultérieurement dévoilée.

L'utilisation de tel ou tel outil dépendra directement de la formulation de la question, de la nature des populations statistiques interrogées, ainsi que du raisonnement engagé dans la comparaison des données que l'observateur aura pu expérimenter. Dans tous les cas, de cette étape d'analyse statistique globale des données, il sera toujours utile de revenir à l'exploitation de comparaisons uni- ou bivariées, plus simples, et qui peuvent s'avérer au final tout aussi informatives par rapport à la question posée (comme le montre l'exemple sur les céramiques, *cf.* 2.1.6). Ce type de comparaison sera également efficace pour décrire avec le plus de précision possible les marqueurs chimiques et/ou géologiques qui semblent être engagés dans les rapprochements ainsi que dans les différences entre les signatures. Il faut donc aussi considérer les analyses statistiques comme la possibilité d'explorer ces jeux de données, du point de vue de l'ensemble des relations qu'elles entretiennent les unes avec les autres. Il faut donc noter la nature fluide des interprétations proposées, à l'image de la représentativité des populations.

5 La relation avec le spatial

Dans les études de provenance, l'analyse cible les relations matérielles entre objets afin d'émettre des hypothèses sur les liens entre des lieux, et éventuellement entre des étapes de transformation. C'est ensuite l'interprétation historique qui permet la discussion sur les mobilités et la diffusion des artefacts voire des procédés, c'est-à-dire

lorsque les relations entre les sites sont rapportées à un espace géographique dessiné par des axes de communication possibles et des caractéristiques physiques (ou culturelles) du milieu. Si le travail de cartographie des différents individus et donc des lieux, de la source à l'usage, est systématique, la relation avec le spatial est en réalité indirecte pour les études de provenance. Autrement dit, les données qui relèvent de la dimension spatiale ne sont pas mobilisées dans le schéma statistique, à l'inverse des exemples de la partie 2.2 qui cherchent, eux, à localiser les objets dans un environnement spatial. Il n'en reste pas moins que les données propres aux formes d'occupation du territoire interagissent dans les raisonnements archéologiques des études de provenance en amont, en apportant une image des occurrences possibles (les entrées), et en aval, en proposant un support cartographique pour les données statistiques afin d'éprouver dans une trajectoire historique les hypothèses relationnelles suggérées (les sorties). Les informations archéologiques disponibles par ailleurs sur la dimension spatiale (répartition géographique, distance entre lieux, caractéristiques physiques du milieu) sont donc contextualisantes aux cas d'étude et intégrées *a posteriori* au schéma statistique. Les méthodes statistiques employées dans les exemples d'étude de cette partie ne permettent donc pas l'analyse des interactions spatiales. Si les données de la dimension spatiale ne sont pas (encore) mobilisées dans les méthodes statistiques des études de provenance, on peut penser que les mettre en interaction avec l'information chimique peut être un moyen d'éprouver plus étroitement les relations que les formes d'organisation des productions du passé entretiennent avec les environnements sociaux et naturels dans une géographie où elles ont pu évoluer. L'observation des relations pourrait alors devenir celle des interactions. Mobiliser les éléments de la dimension spatiale et temporelle (amplitude, changement) peut également être un moyen d'analyser les relations dans un rapport pondérant aux éléments d'un état des connaissances, qui rappelons-le, reste partiel pour les contextes de la majorité des études de provenance.

6 Pour résumer

On comprendra que la méthode statistique miracle pour les études de provenance n'existe pas, et que la recherche relève plutôt d'une démarche progressive, adaptée aux circonstances et dans le cadre évolutif de la recherche. Ce sera également la connaissance de ce que l'on a construit qui permettra d'orienter favorablement les choix qui sont faits. La qualité des réponses données aux questions posées dépendra du schéma analytique choisi, du juste rapport entre information historique et représentativité des données, mais aussi du temps passé à éprouver les jeux de données, c'est-à-dire à en saisir leurs apports et leurs limites dans le travail d'interprétation. C'est aussi vis-à-vis d'une expérience éprouvée dans la durée que l'observateur peut devenir sensible à une particularité. L'accès à l'analyse statistique est également un moyen de tester des interrogations. Prendre le temps de délimiter les contours de la problématique, de formuler la ou les questions de mise en relation avec les sources, d'identifier les lieux et le type de vestiges considérés utiles, de comprendre que l'analyse de comparaison opère dans le référentiel construit à partir des observables qui l'alimentent – et que l'observateur aura donc choisi – semblent être les étapes les plus importantes pour une interprétation prudente et donc opérante.

En résumé :

- Efforcez-vous de formaliser les contours de votre problématique afin de mieux décomposer les composantes propres à la mise en relation pour la recherche de l'origine.
- Estimer la cohérence entre la représentativité des jeux de données et leur signification dans l'analyse vis-à-vis de la question qui est posée.
- Prenez le temps d'éprouver vos jeux de données afin de comprendre ce qui est significatif ou non dans l'analyse statistique (par exemple, des mesures en limite de détection peuvent induire des variations non significatives).
- La pratique joue un rôle important dans la mise en œuvre de l'analyse statistique car elle permet de mettre en interaction l'ensemble des choix qui sont opérés.
- N'hésitez pas à discuter avec des spécialistes qui connaissent les limites d'interprétation, en attendant de devenir vous même spécialiste…