

Les vicissitudes de l'innovation méthodologique

**« Validité, falsifiabilité, parcimonie,
consistance, précision, etc. ¹ »**

Avec environ cent dix publications² depuis les années 1980 et une progression accélérée depuis le début des années 2000, l'analyse de séquences (« AS » dans la suite de ce chapitre) en sciences sociales est une méthode récente. La rapidité de son renouvellement et du développement de ses applications laisse penser qu'elle est loin d'être stabilisée, contrairement par exemple à l'analyse de séries temporelles ou les modèles de durée, qu'elle complète – et concurrence à l'occasion. Elle est un cas stimulant d'histoire scientifique courte, mais intense. Sa cohérence tient à un objet original (les populations de séquences), une façon de le traiter définie par un petit nombre de textes pionniers et un groupe d'utilisateurs s'y consacrant de manière privilégiée. Dans ce texte, j'essaierai de montrer que l'année 2000 fut une date charnière pour la reconnaissance de l'AS comme méthode sociologique, puis que le courant de recherche se développe selon une dialectique entre forces centripètes et centrifuges. Un clivage structurant pour les statistiques sociologiques s'incarne dans ce courant, mais de manière peu conflictuelle, donc sans restreindre la capacité de l'AS à surmonter ses limites actuelles.

1. Levine, 2000, p. 35. Sauf mention contraire, toutes les citations sont traduites par mes soins.
2. Ce total inclut les textes et publications discutant la méthode ou son application à un objet ayant une dimension sociologique, publiés en version papier ou électronique entre le milieu des années 1980 et 2012, quelle que soit la discipline de rattachement des auteurs et le support de publication.

Une dernière défense, puis la révérence : le tournant de 2000

En 2000 paraît dans *Sociological Methods and Research*, l'une des revues qui ont le plus contribué à la reconnaissance de l'AS, une interpellation de Joel Levine à l'attention d'Andrew Abbott. Sous le titre « Qu'avez-vous fait pour nous dernièrement? », Levine commente un article d'Abbott et d'Angela Tsay dans le même volume qui passe en revue les acquis et les développements potentiels de cette méthode. La critique de Levine marque un tournant dans le développement de l'AS : elle formalise et cristallise pour la décennie qui s'ouvre aussi bien les réticences d'une partie de la communauté sociologique que les arguments en sa faveur. Elle est représentative de controverses vives, sur cinq points principaux : la nature des séquences, le rôle de la quantification, le statut épistémologique de la méthode, les coûts de l'algorithme d'appariement optimal (« AO » par la suite) et les modalités de publication³. La réponse d'Abbott (2000) constituera sa dernière contribution écrite substantielle au développement de l'AS à ce jour. Il y récuse de manière assez catégorique les trois premières de ces critiques et répond plus évasivement aux deux suivantes. Douze ans après, les cinq méritent un réexamen car elles font encore l'objet de discussions et tout utilisateur de l'AS s'y trouve confronté tôt ou tard.

Tout d'abord, pour Levine, les séquences sociales sont très différentes des séquences biologiques. Elles n'ont pas de logique génétique et leurs composants n'ont pas la même objectivité. Ce point ne contredit pas Abbott, qui rappelle que le « positivisme narratif » dont il se réclame ne formule pas d'hypothèse sur une logique sous-jacente au déroulement des trajectoires sociales. Biographies, processus historiques ou microsociologiques, les trajectoires sociales laissent toujours une place à l'interaction située et au contexte, qui infléchissent les facteurs structurels et les tendances lourdes. Le développement de l'AS consistera en bonne partie à adapter les outils des biologistes et des informaticiens à cette spécificité des sciences sociales, mais aussi à d'autres : double

3. Le commentaire de Lawrence Wu dans le même volume recoupe celui de Levine à propos des trois dernières critiques de ce dernier. Sauf exceptions qui seront citées le cas échéant au fil du développement, les autres arguments de Wu reposent sur une comparaison approfondie avec les modèles de durée (*event history analysis*) qui ne sera pas reprise ici faute de place.

structure des séquences en états et événements, alphabets des états plus longs et plus complexes, distribution des états plus hétérogène, valeurs manquantes parfois nombreuses, mais porteuses d'une information utile, nécessité de visualiser plus ou moins finement les séquences, etc. L'importation de l'AS, processus toujours en cours, y compris sous forme de nouvelles collaborations avec les biologistes (voir par exemple : Gauthier *et al.*, 2009), est plus qu'une simple transposition.

La seconde critique trouve une réponse tout aussi claire. Levine affirme, citant les erreurs commises avec la *path analysis* ou la régression linéaire, qu'il n'y a pas de garantie de rigueur dans la quantification proposée par Abbott. Or celui-ci n'a jamais insisté sur la nécessité de traiter des données numériques (la plupart des alphabets qui composent les trajectoires sociales sont nominaux) ni de développer des modèles numériques complexes (l'AS est gourmande en mémoire-machine, mais pas plus sophistiquée mathématiquement que les modèles multivariés usuels). Les quatre principales applications qu'il publie (Abbott et Forrest, 1986, Abbott et Hrycak, 1990 ; Abbott et DeViney, 1992 ; Abbott et Barman, 1997) reposent bien plus sur la combinaison pragmatique d'une part de traitements statistiques et graphiques variés et d'autre part d'interprétations sociologiques approfondies. La séduction du chiffre n'est pas son fait, même si, nous le verrons plus loin, elle guette parfois ses continuateurs.

Troisièmement, Levine considère que l'AS est une démarche, pas un modèle, et donc que sa justification exige des épreuves nombreuses et variées. Abbott confirme également ce point. Son projet se démarque depuis les origines (Abbott, 1988, 1992, 1997) du fétichisme des modèles causaux sophistiqués qui, à son sens, a dramatiquement restreint l'horizon théorique de la sociologie américaine depuis les années 1940 et 1950. En particulier, la théorie de l'erreur normale, au fondement de nombre de ces modèles, est souvent faible ou inadaptée aux problématiques sociologiques. Inversement, la statistique descriptive, souvent délaissée, est utile et noble en soi. Découvrir est aussi important que valider. De plus, Abbott présente l'AO, pivot statistique de la grande majorité des études basées sur l'AS, comme « agnostique » sur bien des points, par exemple la définition du temps. Loin de s'enfermer dans une perspective théorique ou ontologique particulière, l'AS est bien plus un outil permettant de rendre compte de manière puissante mais souple de nombre de phénomènes jusqu'alors mal traités ou simplement ignorés.

Dans un quatrième temps, Levine affirme que, si des jeux de coûts différents conviennent également à l'AO, c'est un signe de faiblesse et non de robustesse, contrairement à ce qu'Abbott écrit en 2000 (et déjà dans Abbott et Hrycak, 1990, p. 163-164). Depuis les années 1990, c'est la question des coûts qui a été, à tort ou à raison, l'aspect le plus discuté de l'AS. Effectivement, si toutes les options se valent, la matrice de substitution anti-identité⁴ devrait s'imposer, couplée avec des coûts d'insertion-suppression fixés à environ 0,5. Or la majorité des auteurs refuse cette solution, au profit de formules spécifiques aux données traitées. La conclusion d'Abbott semble donc prématurée et dépendante des données qu'il a eu l'occasion de traiter. La section 4 reviendra sur la question et, plus généralement, sur la validation de la méthode et son degré de généralité.

Dans sa cinquième critique, Levine (comme Wu) réclame que le détail des données et des équations soit systématiquement publié pour favoriser la réplication et les tests. Abbott répond que la réplication en sciences sociales devrait être souple, que le hasard y a sa place. Il défend implicitement une position dualiste : le modèle des sciences de la nature est inadapté aux terrains mouvants et éphémères des observateurs de la société humaine. Il détaille autant que possible les éléments mathématiques de ses démonstrations, mais sans leur subordonner, comme cela se trouve parfois, la présentation littéraire de son raisonnement. Avec le recul apparaît le poids de contraintes extrascientifiques – éditoriales, financières et organisationnelles. Bien que le travail scientifique ait à présent les moyens techniques de s'émanciper du support papier, les usages continuent à s'appuyer essentiellement sur le texte principal. La question de l'administration de la preuve en matière d'AS, en matière de transparence des exposés, mais aussi de complétude des démonstrations, reste donc ouverte, elle aussi.

En substance, Levine rappelle les exigences méthodologiques incontournables mentionnées dans le titre du présent chapitre. Abbott serait allé trop vite, omettant d'asseoir l'AS auprès des initiés avant de la diffuser plus largement. L'intéressé répond que l'AS est encore imparfaite, mais que rien ne la remplace sur certaines données. Les résultats déjà accumulés montrent qu'elle est déjà aussi utile que des méthodes

4. La matrice faite de 0 sur la diagonale principale et de 1 ailleurs n'a pas de nom spécifique, elle peut être dénommée anti-identité par référence à la matrice identité, faite de 1 sur la diagonale et de 0 ailleurs.

établies qui accommodent tout autant partis pris théoriques, outils statistiques et données disponibles. La force de conviction des modèles de régression et autres modèles de durée repose certes sur des fondations mathématiques, mais également sur des routines de raisonnement qui clôturent la discussion et masquent leurs limites. Les conditions d'application du modèle linéaire général (Abbott, 1988) sont très restrictives et souvent inadaptées aux sciences sociales. La force de l'AS est de lever ces contraintes et de dégager des horizons pour l'imagination théorique. La dispute de 2000 révèle les vicissitudes de l'innovation méthodologique et les résistances de la science normale (Kuhn, 1962) aux idées nouvelles.

La légitimation par la standardisation

À ce jour, l'AS a gagné en légitimité institutionnelle et se normalise, mais les interrogations de ses débuts restent ouvertes. Le courant de recherche est pris entre deux tendances inverses, l'une centripète (développée dans la présente section) et l'autre centrifuge (voir *infra*, section suivante). La logique centripète manifeste un projet collectif de standardisation de la méthode. Cette ambition est portée par un collège d'utilisateurs soudé, qui partagent la connaissance des textes initiaux, notamment ceux d'Abbott et de ses collègues, et se lisent les uns les autres. Ils font converger les efforts des uns et des autres pour légitimer la méthode et accéder à une cumulativité des principes d'analyse, des procédures et des résultats. À l'issue de la période des pionniers des années 1980 et 1990⁵, Abbott ayant renoncé au rôle de chef de file pour se consacrer à d'autres recherches, des standards s'imposent peu à peu.

Ainsi, la quasi-totalité des auteurs entérine la centralité de l'algorithme d'AO comme outil de réduction de la diversité des profils de séquences. Rappelons⁶ que l'AO compare deux séquences S1 et S2 en appliquant successivement à chacun des états composant S1 des opérations élémentaires, par exemple l'insertion et la suppression d'états, de

5. Voir, dans ce volume, le chapitre de Nicolas Robette (« Du prosélytisme à la sécularisation. Le processus de diffusion de l'«*optimal matching analysis*» », *infra*).
6. Pour un rappel plus détaillé sur l'analyse de séquences, voir par exemple: Blanchard, 2011; Lesnard, 2006; voir aussi, dans le vol. 1 de cet ouvrage, le texte de Heather MacIndoe et Andrew Abbott, « Analyse de séquences et techniques d'appariement optimal pour les sciences sociales », trad. par Claire Lemercier et Nicolas Robette, p. 357.

manière à obtenir S2. La dissimilarité entre S1 et S2 est le décompte de ces opérations élémentaires, éventuellement pondérées, par exemple par la nature des états rencontrés. Finalement, plusieurs séquences d'opérations élémentaires étant en général envisageables, on retient la plus « économique ». Il est à noter qu'Abbott ne considérait pas l'AO comme une solution exclusive, anticipant même plutôt un foisonnement de propositions alternatives et concurrentes⁷. Or, les propositions de calculs alternatifs des dissimilarités par paires (Dijkstra et Taris, 1995 ; Elzinga, 2003 et 2006) sont peu suivies. De même, la modélisation intégrale des trajectoires (Abell, 1987), l'apprentissage par intelligence artificielle (Billari *et al.*, 2000) ou le traitement des séquences par l'analyse de correspondance (Robette et Thibault, 2008) restent des tentatives isolées, bien que prometteuses. La simplicité et la robustesse de l'AO s'imposent auprès des utilisateurs non spécialistes de la combinatoire séquentielle et de la programmation informatique.

De fait, les résultats encourageants obtenus par les premières applications ont entraîné l'adhésion de nombreux auteurs à trois principes : la comparaison incrémentale systématique de S1 et S2, c'est-à-dire état par état dans le sens du temps, de préférence au repérage automatisé de blocs communs à S1 et S2 ; l'alignement, c'est-à-dire la troncature du temps visant à repérer des sous-séquences identiques (motifs ou *patterns*) localisées à différents endroits des séquences ; la sélection systématique du plus court des chemins de comparaison entre S1 et S2 au moyen de l'algorithme de Saul Needleman et Christian Wunsch, au détriment d'autres chemins possibles. La cristallisation informatique de ces choix par l'AO contribuera à leur reconduction, d'*Optimize* (Abbott et Prellwitz, 1997) à *Transition Data Analysis* (Rohwer et Poetter, 2007), quoique *Traminer* (Gabadinho *et al.*, 2011) propose des alternatives.

Des paramètres standards complémentaires sont fixés à l'intérieur de l'AO, encore une fois de manière sélective. Les opérations élémentaires utilisées pour transformer une séquence en une autre et calculer l'indice de dissimilarité sont réduites à quatre : l'insertion, la suppression, la substitution et la correspondance (*match*). L'opération de permutation (*swap*) est par exemple écartée, sans débat notable. De la même manière, une majorité d'études reconduit une série de cinq conventions sur les coûts : insertion et suppression sont inverses et de coût égal ; leur coût est fixe,

7. Commentaire d'Abbott lors du colloque « Les parcours sociologiques d'Andrew Abbott » (université de Versailles Saint-Quentin-en-Yvelines, mai 2011).

alors que certains ont proposé de les faire varier dans le temps (Lesnard, 2010) ou suivant les états concernés ; les coûts de substitution sont symétriques, alors que la probabilité empirique ou théorique du passage d'un état A à un état B n'est pas toujours la même que de B à A, pas plus que le « coût » social que l'on peut assigner à l'un et l'autre ; les coûts de substitution de A à B sont fixes dans le temps, quand on pourrait les faire varier suivant les voisins de A et B dans leurs séquences empiriques respectives (Hollister, 2009) ; enfin, l'utilisation d'informations empiriques (coûts de transition) ou théoriques (coûts objectifs) relatives aux transitions individuelles dans le temps *réel* pour établir les coûts de la comparaison des séquences dans le temps *tronqué* de l'AO, bien que signalée discrètement par Abbott comme problématique (Abbott et Hrycak, 1990, p. 162-163), n'est pas discutée.

Dernier aspect du processus de standardisation, la stratégie de traitement des séquences suit souvent les étapes consacrées suivantes : AO, classification, interprétation, plus une phase de description graphique et statistique des séquences brutes ou après classification. Le succès de cette séquence de traitement éclipse des aspects pourtant intéressants des articles des années 1990 : le codage de matériaux originaux et complexes (Abbott et Barman, 1997, p. 49-60) ; la comparaison des séquences réalisées avec leur probabilité attendue sur l'ensemble de la population (Abbott et DeViney, 1992, p. 263-265) ; la standardisation ou la transformation logarithmique des durées des séquences (Abbott et Hrycak, 1990, p. 169-170 ; Abbott et DeViney, 1992, p. 271) ; les stratégies mixtes de détermination des coûts (par exemple les coûts objectifs ajustés au moyen de coûts de transition : *ibid.*, p. 162-163) ; ou encore la combinaison de l'AS avec un modèle de durée (Stovel, 2001, p. 866). Mais les logiques de rentabilité propre à la production scientifique et à la sélection éditoriale, ajoutées à l'entreprise collective de légitimation de la méthode, réduisent la prise de risque et les innovations.

Maintenir néanmoins la biodiversité disciplinaire

À dire vrai, en plus du ralliement à la série de conventions méthodologiques exposées ci-dessus, le succès de la voie standard s'explique par son adaptation particulière aux applications empiriques dominantes.

Plus de 80 % de la littérature est occupée par la sociologie des carrières professionnelles, qui a convergé au cours des années 2000 avec la sociologie des parcours familiaux, de loisirs et résidentiels dans l'approche dite des « parcours de vie ». Il s'agit en général d'étudier des trajectoires individuelles, à l'échelle de l'année, et en particulier les périodes de transition, souvent critiques, entre jeunesse et vie adulte ou entre vie active et troisième âge, le tout dans une perspective contemporaine. Abbott, bien que sociologue du travail, n'y voyait pourtant qu'une application parmi d'autres. L'examen dans cette section de deux domaines d'application minoritaires, la sociologie historique et la sociologie politique⁸, révèle en creux la réduction opérée par le processus de standardisation. Il met aussi en évidence le profit que la méthode peut tirer du métissage entre ces domaines.

La sociologie historique est naturellement intéressée par le regard neuf que l'AS pose sur les temporalités. L'étude des épisodes d'émeutes raciales dans le sud-est des États-Unis de 1882 à 1930 (Stovel, 2001) montre qu'il est possible de recenser systématiquement des événements rares, d'en déduire des cycles de mobilisation interprétables par des contextes sociaux transversaux aux *counties* où ils se produisent. Claire Lemercier (2005) s'intéresse, quant à elle, aux carrières des élites consulaires françaises au XIX^e siècle. Elle montre comment certains individus multipositionnés font le lien entre les institutions économiques, tandis que d'autres occupent des postes plus isolés et plus brefs. Enfin, Fabien Accominotti (2009) reconstitue les carrières de créativité des peintres actifs en France nés entre 1830 et 1900 sur la base de leur visibilité dans les manuels d'histoire de l'art anglais et américains. Il teste ainsi l'idée d'opposition entre créateurs précoces et tardifs, et montre comment certains courants artistiques, en particulier l'impressionnisme, déterminent les possibles professionnels et façonnent les carrières individuelles. Dans ces deux derniers exemples, la sociologie historique renouvelle l'approche par les carrières des sociologues du contemporain.

L'intérêt pour l'historicité des séquences étudiées, par contraste avec les séquences des démographes réduites aux logiques intra-individuelles, rapproche les politistes des historiens. Mais les séquences politiques,

8. L'économie, la géographie ou l'ethnographie font également figure de domaines d'application mineurs. Pour une revue de littérature plus systématique, on pourra lire: Abbott, 1995, Abbott et Tsay, 2000, et pour les parcours de vie Aisenbrey et Fasang, 2010.

échues ou contemporaines, sont marquées aussi par les phénomènes de compétition et de coopération, qu'il s'agisse de la conquête et de l'exercice du pouvoir, de la mise en place d'une politique publique ou de l'expression d'une idéologie. Les interactions entre individus et contextes de déploiement des trajectoires pèsent donc particulièrement. C'est par exemple visible lorsque l'AS est appliquée aux processus d'adoption de législations sociales dans dix-huit pays industrialisés sur un siècle (Abbott et DeViney, 1992), intriquant temporalités nationales (hypothèse du développement autonome de l'État-providence), internationales (hypothèse de la diffusion) et transnationales (hypothèse du processus mondial). L'AO prend ici en compte non seulement l'ordre des événements, mais aussi le moment historique où surviennent des avancées législatives. Les régimes politiques peuvent aussi être étudiés à travers les séquences de crise, sur la base d'une échelle standardisée de conflit et de coopération par pays (Casper et Wilson, 2015). Le déclenchement et la résorption des conflits apparaissent liés aux caractéristiques des régimes, aux acteurs impliqués et au type d'interactions entre eux. Par contraste, dans des situations plus routinisées, les trajectoires politiques dépendent certes des ressources sociales des acteurs et de leurs interactions, mais elles sont aussi cadrées par des règles, formelles (lois, règlements) et informelles (normes morales).

Deux derniers objets politiques témoignent des potentialités de l'AS au-delà des parcours de vie : le vote et le militantisme. Les études électorales tirent ainsi parti de l'analyse systématique des trajectoires de vote. Si l'on fait abstraction des contextes historiques, des interactions micro-sociales imprévues peuvent apparaître (Buton *et al.*, 2012). Mesurées par le fait de participer ou non aux scrutins successifs, ces trajectoires font ressortir l'alignement des membres d'un même foyer les uns sur les autres, justifiant l'examen de la maisonnée comme lieu de politisation à part entière. Versant non conventionnel mais tout aussi déterminant de la participation politique que le vote, le militantisme relève de l'engagement individuel au profit d'une cause qui dépasse l'individu. Camille Masclat (recherche en cours) s'intéresse ainsi aux effets des trajectoires d'engagement des féministes sur la vie professionnelle, intime et militante de leurs filles. Elle montre combien investissements privés et publics peuvent être liés, temps biographique et temps politique s'articulant à travers des expériences vécues ensemble. Il en va de même pour les membres d'une association militante (*Aides*), qui présentent des profils biographiques similaires suivant les générations auxquelles ils appartiennent

(Blanchard, 2010 ; Fillieule et Blanchard, 2013). Les temporalités individuelles s'emboîtent dans celles de l'organisation et, au-delà, dans celles de la mouvance militante concernée et de l'espace sociopolitique global.

Sociologie historique et sociologie politique présentent donc des variations stimulantes aux études des parcours de vie. Les emboîtements de temporalités en sont un exemple, avec des conséquences notables sur les stratégies de codage, de représentation graphique et de calcul des dissimilarités entre séquences. Le problème de la nature des séquences évoqué par Levine est également revisité, car les séquences historiques et politiques divergent du modèle génétique plus encore que les parcours de vie : l'accent mis sur les interactions situées réduit plus encore le poids de déterminations structurelles. Au-delà, les enjeux historiques et politiques posent de nouvelles questions sur l'entremêlement des temps biographique, organisationnel et historique. Il s'ensuit que certains éléments de l'alphabet peuvent être définis par des événements organisationnels, comme la création de l'organisation étudiée (Blanchard, 2010), ou que les graphes séquentiels individuels peuvent être indexés sur une césure individuelle-organisationnelle, comme l'entrée dans l'organisation (*ibid.* ; Colombi et Paye, 2014).

Une intrication originale entre deux traditions statistiques

La critique de Levine révèle un autre clivage, transversal aux spécificités disciplinaires, entre deux manières de pratiquer les statistiques. En première approximation, qualifions-les de traditions modélisatrice et descriptive. Ces deux manières, présentes dans la statistique appliquée depuis le XVIII^e siècle (Desrosières, 1993), sont des assemblages stables, bien que sans coordination forte, d'objectifs de recherche, de théories, de méthodes, d'outils et de croyances ontologiques et idéologiques, qui soudent une communauté de savants et s'incarnent dans des textes et des applications de référence. À ce titre, elles se rapprochent de « paradigmes » (Kuhn, 1962) ou plus précisément de « programmes de recherche » – voir Lakatos et Musgrave, 1970 – qui coexistent, se livrent concurrence et conquièrent chacun plus ou moins de terrain suivant les époques. Ces programmes se révèlent dans la recherche sur les séquences quoique les auteurs ne se rattachent pas tous strictement à l'un ou l'autre. L'objectif

de contribuer de manière cumulative à améliorer la compréhension des phénomènes séquentiels l'a jusqu'à présent emporté sur la volonté d'expliquer son affiliation à un programme (en dépit des avantages qu'une telle stratégie apporte en matière de sécurisation des débouchés éditoriaux et des parcours institutionnels).

La tradition modélisatrice met l'accent sur l'élaboration de modèles statistiques causaux. L'idéal implicite des sciences physiques oriente ses représentants vers la formalisation informatique, la formulation quantifiée des résultats, la précision des décomptes et des mesures, la possibilité de réplication et l'inférence d'échantillonnage. Les critères de scientificité évoqués par Levine (2000) s'y rattachent. Abbott (1988), bien que dénonçant la réduction opérée par l'obsession modélisatrice du modèle linéaire général, s'y réfère dans une certaine mesure. Son projet de positivisme narratif visait à combler les lacunes de l'approche par les récits de vie et les études de cas historiques.

Ce courant utilise prioritairement les grandes bases de données secondaires, panels individuels et familiaux, recensements nationaux, suivis de cohortes d'étudiants, etc. (par exemple : Bühlmann, 2010 ; Gauthier, 2007 ; Lesnard, 2010 et Robette, 2010). Les présupposés et les limites de ces données sont moins discutés que leur traitement. Afin de mettre au point des tests de robustesse, cette tradition peut recourir à des données simulées (Halpin, 2010 ; Robette et Bry, 2012 ; Studer *et al.*, 2011), en particulier en introduisant du bruit, c'est-à-dire des valeurs manquantes (Gauthier, 2007 ; Casper et Wilson, 2015). Elle tend à concentrer son attention sur l'amélioration ou la critique de l'algorithme d'AO (Bison, 2009 ; Elzinga, 2006 ; Hollister, 2009). Elle défend une application rigoureuse de celui-ci, par exemple en respectant strictement l'inégalité triangulaire entre les coûts de substitution, en ne traitant que les séquences de longueur homogène et en excluant les séquences trop fournies en valeurs manquantes. Elle résume aussi volontiers des données de séquences nominales au moyen d'indicateurs continus de complexité, de volatilité ou d'entropie, qui mesurent par exemple la (dé)standardisation des parcours de vie (Elzinga, 2010 ; Widmer *et al.*, 2003). Elle vise le rattachement de l'AS à l'approche individus-variables, par exemple à travers la *discrepancy analysis*, méthode qui évite la part d'arbitraire introduite par la procédure de classification en liant plus directement la matrice des distances aux covariables (Studer *et al.*, 2011). En définitive, cette tradition met l'accent sur la consistance formelle de l'AS, comme condition de l'administration de la preuve et

de la généralisation de la méthode. C'est elle qui contribue le plus à la cohérence du collège des analystes de séquences depuis une décennie.

À l'inverse, la tradition descriptive met l'accent sur l'élaboration théorique, la stimulation de l'imagination sociologique et la pertinence des résultats, plus que sur la rigueur de l'administration de la preuve et sa valeur universelle. Elle accorde un espace notable à la justification de la pertinence théorique des données, à leur exploration et à leur compréhension par la multiplication des représentations numériques et graphiques. Elle valide, explicitement ou non, les critiques du modèle linéaire général formulées par Abbott : le monde social est moins constitué d'individus définis comme conjonctions de propriétés mesurées par les variables que comme des individus qui vivent des expériences successives dans des contextes donnés (Abbott et Tsay, 2000, p. 24). Cette tradition recourt souvent à des échantillons moins volumineux que la précédente. Quelques dizaines d'individus suffisent (Abbott et Forrest, 1986 ; Casper et Wilson, 2015). Andrew Abbott et Stanley DeViney (1992) travaillent sur une population quasi complète : pour eux, la question de la représentativité ne se pose guère. Les séquences sont élaborées de manière artisanale, loin des routines des grandes enquêtes standardisées. Leur nature s'écarte logiquement du modèle biographique : processus historiques à l'échelle d'un pays (Abbott et DeViney, 1992 ; Wilson, 2014), productions symboliques (Abbott et Forrest, 1986 ; Abbott et Barman, 1997), séquences de déplacements quotidiens (Wilson, 1998 ; Schlich, 2003), pratiques ritualisées (Abbott et Forrest, 1986 ; Forrest et Abbott, 1990), pratiques formalisées (Buton *et al.*, 2012) ou micro-interactions sociales (Kreuter et Kohler, 2009 ; Pollien et Joye, 2014). Les temporalités se diversifient aussi, en matière d'échelles et d'unités de temps. L'axe temporel peut même reposer sur des événements irréguliers (Abbott et Forrest, 1986 ; Buton *et al.*, 2012) ou des coups échangés (Casper et Wilson, 2015) et donc ne pas être normé : ce sont l'ordre des événements et la densité sociale du temps (Durkheim, 1990) qui priment alors.

Ces variations provoquées par la tradition descriptive renouvellent les questions sur le statut de l'AO et sur ses paramétrages : que mesurent les dissimilarités ? Quel sens prend la transposition, problématique (voir section 2), d'informations diachroniques sur une comparaison atemporelle des séquences ? Devant ces questions difficiles, il semble, bien que ce point soit rarement explicité, que certains auteurs utilisent l'AO comme une technique en partie aveugle mais efficace. La représentation

en graphes de trajectoires individuelles de la classification issue de la matrice des distances peut leur suffire à justifier la méthode : elle introduit un ordre auparavant invisible dans le chaos de l'échantillon d'origine. Cette perspective pragmatique s'oppose complètement aux objectifs de formalisation propres à la tradition modélisatrice. Dans le même ordre d'idées, le travail sur données primaires pousse à relativiser la perfection formelle de l'outil statistique. Les données « faibles » (valeurs manquantes, ambiguës, incertaines), au lieu d'être exclues de la collecte des données, peuvent être conservées telles quelles jusqu'à l'interprétation des résultats. Cette stratégie inclusive génère par exemple des classes de séquences marquées à divers degrés par les conditions de passation et de codage de l'enquête, bien identifiées comme telles, qui, loin d'appauvrir le résultat d'ensemble, l'enrichissent (Blanchard, 2010). Il en va de même pour les séquences censurées : étudier les trajectoires organisationnelles en excluant respectivement du corpus les individus mal renseignés à gauche (pour des problèmes de mémoire par exemple) ou à droite (personnes dont la trajectoire organisationnelle est encore en cours au moment de l'enquête) fausse la compréhension du phénomène, puisque les queues de distribution contribuent aux interactions dans l'organisation et à la dynamique collective. Les garder sauvegarde la diversité des profils. Les plus censurés sont recueillis par des classes spécifiques, caractérisées, ici aussi, à la fois par le matériau sociologique et par les conditions de passation de l'enquête. Les imperfections de l'enquête sont maintenues, mais contrôlées.

Autre conséquence du pragmatisme formel de la tradition descriptive, ses partisans subordonnent moins strictement la matrice des coûts à la règle de l'inégalité triangulaire. Cette règle, qui suppose d'associer une métrique parfaite à l'espace des états qui composent les séquences, est rendue difficile par la pluralité des interprétations sociologiques. Des entorses à la règle sont donc tolérées, si elles permettent un gain de signification. Dans le même ordre d'idées, il est possible de traiter des séquences de longueurs inégales, dues par exemple à des durées de vie inégales, en tirant parti du rôle d'ajustement assuré par les insertions-suppressions (Blanchard, 2010). C'est souvent mutiler la réalité sociale que de délimiter des tronçons biographiques par des âges ou des dates qui ne correspondent pas aux logiques sociales à l'œuvre. Au contraire, on mise sur la simplicité de formulation et la robustesse de l'AO face à des données hétérogènes. Les variations de paramétrage possibles n'apparaissent pas comme des faiblesses à corriger (voir Levine, 2000) mais comme des

moyens de moduler la dimension du temps visée par l'analyse. La proportion entre les différents coûts permet de mettre l'accent sur la durée, l'ordre ou la position des états sur l'axe du temps, en fonction du problème de recherche (Lesnard et de Saint Pol, 2006 ; Blanchard, 2010). Considérons par exemple les trajectoires d'un échantillon aléatoire de Français du point de vue de leurs votes, année après année, avec quatre modalités : *abstention ou non-inscrit, gauche, droite et autre*. La longueur des séquences est notablement hétérogène du fait des âges variés des répondants. Si les insertions et suppressions, seul moyen d'ajuster deux séquences de longueur différente, se voient attribuer un coût faible, alors les différences d'âge, c'est-à-dire la durée des séquences, sont minorées. L'AO discrimine donc les séquences avant tout du point de vue de la nature et de l'ordre des états successifs. Il contraste des séquences stables de nature différente (par exemple un abstentionniste régulier et un votant de droite fidèle), mais aussi des séquences qui varient différemment les unes des autres (par exemple un votant de gauche déçu passant à l'abstention ou un votant de droite déçu passant à gauche)⁹.

Enfin, la tradition descriptive voit dans l'AO un élément de la boîte à outils qu'est l'AS, laquelle offre une diversité de traitements avant et après l'AO. Elle fait notamment un usage abondant des graphes séquentiels individuels et agrégés (Lesnard, 2010), y compris multiséquentiels (Fillieule et Blanchard, 2013), ainsi que des cartes basées sur le *multidimensional scaling* (MDS) (Piccarreta, Lior, 2010). Ces graphes sont rendus possibles par les capacités nouvelles des ordinateurs et les programmations *ad hoc*, comme la librairie *Traminer* qui tire profit de la puissance graphique du logiciel *R*. L'explication sociologique rompt ainsi avec le schéma traditionnel du modèle de régression résumé en un grand tableau à multiples décimales et un petit nombre de statistiques de test définitives, tranchant péremptoirement un jeu d'hypothèses dichotomiques. Elle découle plutôt de l'usage pondéré et théorisé de différents points de vue sur une population de séquences dont le sens ne s'épuise pas dans des statistiques de résumé.

Le cas de l'AS est remarquable car les deux approches, modélisatrice et descriptive, coexistent et coopèrent plus qu'elles ne se confrontent.

9. Les proportions entre coûts ont des conséquences plus subtiles qui ne peuvent être détaillées ici. Dans cet exemple, la petitesse des coûts d'insertion-suppression peut différencier plus ou moins deux séquences de même longueur, suivant le rythme des changements de vote.

C'est plutôt la seconde qui explique le succès de l'AS en France, en affinité avec une pratique sociologique qui avait fait bon accueil en son temps à l'analyse de données de Benzécri (1982). La première approche a plus de succès dans une tradition héritée des sciences sociales américaines des années 1950 et 1960 et perpétuée depuis lors en Europe, bien que les chercheurs ayant promu l'AS aux États-Unis (Andrew Abbott, Silke Aisenbrey, Anette Fasang, Shin-Kap Han, Katherine Stovel) s'inspirent conjointement des deux traditions. Modélisation et description convergent : la seconde utilise (sélectivement) certains des outils statistiques et informatiques mis au point par la première, laquelle emprunte en retour certains des aménagements pragmatiques opérés par la seconde. La place manque ici pour détailler les méandres des confrontations et des coopérations entre deux courants qui portent deux conceptions différentes des sciences sociales.

Pragmatisme n'est pas opportunisme

La diversité disciplinaire et la cohabitation féconde entre deux traditions statistiques promettent des innovations encore importantes pour l'AS. La phase de décollage évoquée en introduction devrait perdurer, au vu des zones qui restent dans l'ombre. Citons-en deux. La première concerne les effets conjugués des différents paramétrages de l'AO. Hormis les implications pratiques des conceptions du temps social (Lesnard, 2010) et les effets d'interaction entre les différents coûts (Bison, 2009), la plupart des tests ont considéré certains paramètres isolément, alors que l'AO est holiste par construction. La seconde zone d'ombre concerne l'assise empirique de l'AS. Les tests sont encore en nombre limité, hormis dans le domaine des parcours de vie. En particulier, la diversité des structures empiriques de populations de séquences n'a été qu'effleurée : comment imaginer que les traitements (AO, MDS, graphes, statistiques de complexité, etc.) puissent s'appliquer de la même manière et avec le même profit à des ensembles de séquences qui diffèrent par la taille des alphabets, par la distribution des états, par la dimensionnalité de l'espace que forment ces états, par la distribution de longueurs des séquences et par le degré de réversibilité des enchaînements entre états ? Les bilans et les méta-analyses (Abbott, 1995 ; Abbott et Tsay, 2000 ; Gauthier, 2007 ;

Aisenbrey et Fasang, 2010) peuvent apporter des solutions, mais la simulation, encore rare, permettrait une exploration plus systématique.

Ces incertitudes expliquent que la justification des choix méthodologiques mêle, en proportion variable d'un auteur à l'autre, quatre types d'arguments: la fondation formelle, de la théorie statistique aux outils informatiques; le pragmatisme empirique; l'optimisation sur un terrain donné de certains paramètres contrôlables, comme les coûts de l'AO; la reproduction des choix éprouvés sur d'autres terrains. Nombre d'analystes de séquences font preuve sur ce quatrième point d'une forme d'opportunisme épistémologique, au nom de la cause commune. Cela ne décrédibilise pas l'AS, mais, faute de s'entendre sur une ligne d'argumentation commune, la cumulativité méthodologique s'en ressent et certains lecteurs habitués aux justifications plus homogènes conservent leurs doutes.

Finalement, où va l'AS? En l'absence d'arbitrage pour recueillir son héritage, suite au retrait anticipé de son fondateur, et du fait du caractère encore artisanal de ce champ, personne ne contrôle le processus de sélection historique et aucun facteur sélectif ne semble dominer. On constate plutôt une combinaison en partie aléatoire de plusieurs d'entre eux: les exigences éditoriales de formats courts et standardisés, qui desservent certains produits atypiques de l'AS; les attentes des experts non formés à l'AS; le nombre limité de débouchés en phase de décollage paradigmatique, qui implique de conquérir et de défendre des lieux de publication et d'enseignement; les frontières disciplinaires, qui limitent les mises en commun nécessaires; les stratégies individualisées de prise de risque dans les soumissions aux revues et la tentation de se conformer aux précédents plutôt que d'innover; la disponibilité des outils informatiques pour lesquels on est passés d'une situation de pénurie avec un coût élevé à l'entrée, à une situation de relative facilité qui pourrait encourager la diversité, ou inversement la réduire. En bref, sans feuille de route rédigée à l'avance, l'AS reste une aventure et non la sinécure que d'autres méthodes ont pu devenir.

Bibliographie

- ABBOTT Andrew, 2001, *Time Matters: On Theory and Method*, Chicago, University of Chicago Press.
- 2000, « Reply to Levine and Wu », *Sociological Methods and Research*, vol. 29, n° 1, p. 65-76.
- 1995, « Sequence analysis: New methods for old ideas », *Annual Review of Sociology*, vol. 21, p. 93-113.
- 1998, « The causal devolution », *Sociological Methods and Research*, vol. 27, n° 2, p. 148-181.
- 1997, « Seven types of ambiguity », *Theory and Society*, vol. 26, nos 2-3, p. 357-391.
- 1992, « From causes to events: Notes on narrative positivism », *Sociological Methods and Research*, vol. 20, n° 4, p. 428-455.
- 1988, « Transcending general linear reality », *Sociological Theory*, vol. 6, n° 2, p. 169-186.
- ABBOTT Andrew et BARMAN Emily, 1997, « Sequence comparison via alignment and Gibbs sampling: A formal analysis of the emergence of the modern sociological article », *Sociological Methodology*, vol. 27, n° 1, p. 47-87.
- ABBOTT Andrew et DEVINEY Stanley, 1992, « The Welfare State as transnational event: Evidence from sequences of policy adoption », *Social Science History*, vol. 16, n° 2, p. 245-274.
- ABBOTT Andrew et FORREST John, 1986, « Optimal matching methods for historical sequences », *Journal of Interdisciplinary History*, vol. 16, n° 3, p. 471-494.
- ABBOTT Andrew et HRYCAK Alexandra, 1990, « Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers », *American Journal of Sociology*, vol. 96, n° 1, p. 144-185.
- ABBOTT Andrew et PRELLWITZ Grant, 1997, *Optimize* [logiciel], New York, Prellwitz Computing Services. Archive de la page de présentation de juin 2011 disponible en ligne : web.archive.org/web/20110607071114/http://home.uchicago.edu/~aabbott/om.html (consulté en mars 2016).
- ABBOTT Andrew et TSAY Angela, 2000, « Sequence analysis and optimal matching methods in sociology: Review and prospect », *Sociological Methods and Research*, vol. 29, n° 1, p. 3-76.
- ABELL Peter, 1987, *The Syntax of Social Life: The Theory and Method of Comparative Narratives*, Oxford, Clarendon Press.
- ACCOMINOTTI Fabien, 2009, « Creativity from interaction: Artistic movements and the creativity careers of modern painters », *Poetics*, vol. 27, n° 3, p. 267-294.

- AISENBREY Silke et FASANG Anette, 2010, «New life for old ideas: The “second wave” of sequence analysis. Bringing the “course” back into the life course», *Sociological Methods and Research*, vol. 38, n° 3, p. 420-462.
- BENZÉCRI Jean-Paul, 1982, *Histoire et préhistoire de l'analyse des données*, Paris, Dunod.
- BILLARI Francesco, FÜRNKRANZ Johannes et PRSKAWETZ Alexia, 2000, «Timing, sequencing and quantum of life course events: A machine learning approach», Rostock, Max-Planck-Institut für demografische Forschung Working Paper 2000-010.
- BISON Ivano, 2009, «OM matters: The interaction effects between *indel* and substitution costs», *Methodological Innovations Online*, vol. 4, n° 2, p. 53-67.
- BLANCHARD Philippe, 2011, «Sequence analysis for political science», *Working Papers of the Committee on Concepts and Methods*, International Political Science Association, en ligne: www.concepts-methods.org/Files/WorkingPaper/PM_32_Blanchard.pdf (consulté en mars 2016).
- 2010, «Analyse séquentielle et carrières militantes», rapport de recherche, en ligne: hal.archives-ouvertes.fr/hal-00476193v2 (consulté en mars 2016).
- BLANCHARD Philippe, BÜHLMANN Felix et GAUTHIER Jacques-Antoine (eds.), 2014, *Advances in Sequence Analysis: Theories, Methods and Applications*, New York, Springer.
- BÜHLMANN Felix, 2010, «Routes into the British service class: Feeder logics according to gender and occupational group», *Sociology*, vol. 44, n° 2, p. 195-212.
- BUTON François, LEMERCIER Claire et MARIOT Nicolas, 2012, «The household effect on electoral participation: A contextual analysis of voter signatures from a French polling station (1982-2007)», *Electoral Studies*, vol. 31, n° 2, p. 434-447.
- CASPER Gretchen et WILSON Matthew, 2015, «Using sequences to model crises», *Political Science Research and Methods*, vol. 3, n° 2, p. 381-397.
- COLOMBI Denis et PAYE Simon, 2014, «Synchronising sequences: An analytic approach to explore relationships between events and temporal patterns», dans Philippe Blanchard, Felix Bühlmann et Jacques-Antoine Gauthier (eds.), *Advances in Sequence Analysis: Theories, Methods and Applications*, New York, Springer.
- DESROSIÈRES Alain, 1993, *La politique des grands nombres. Histoire de la raison statistique*, Paris, La Découverte.
- DIJKSTRA Wil et TARIS Toon, 1995, «Measuring the agreement between sequences», *Sociological Methods and Research*, vol. 24, n° 2, p. 214-231.

- DURKHEIM Émile, 1990, *Les formes élémentaires de la vie religieuse*, Paris, Puf. Édition originale: Paris, Félix Alcan, 1912.
- ELZINGA Cees H., 2010, «Complexity of categorical time series», *Sociological Methods and Research*, vol. 38, n° 3, p. 463-481.
- 2006, «Sequence analysis: Metric representations of categorical time series», rapport de recherche, Department of Social Science Research Methods, Amsterdam, Université libre d'Amsterdam.
- 2003, «Sequence similarity: A nonaligning technique», *Sociological Methods and Research*, vol. 32, n° 1, p. 3-29.
- FILLIEULE Olivier et BLANCHARD Philippe, 2013, «Fighting together: Assessing continuity and change in social movement organisations through the study of constituencies' heterogeneity», dans Kauppi Niilo (ed.) *A Political Sociology of Transnational Europe*, Colchester, ECPR Press, p. 79-108.
- FORREST John et ABBOTT Andrew, 1990, «The optimal matching method for studying anthropological sequence data», *Journal of Quantitative Anthropology*, vol. 2, n° 2, p. 151-170.
- GABADINHO Alexis, RITSCHARD Gilbert, STUDER Matthias et MÜLLER Nicolas S., 2009, *Mining Sequence Data in R with the Traminer Package: A User's Guide*, Département d'économétrie et laboratoire de démographie, Université de Genève.
- GAUTHIER Jacques-Antoine, 2007, *Empirical Categorizations of Social Trajectories: A Sequential View on the Life Course*, thèse de doctorat, Université de Lausanne (dactyl.).
- GAUTHIER Jacques-Antoine, WIDMER Éric, BUCHER Philip et NOTREDAME Cédric, 2009, «How much does it cost? Optimization of costs in sequence analysis of social science data», *Sociological Methods and Research*, vol. 38, n° 1, p. 197-231.
- HALPIN Brendan, 2010, «Optimal matching analysis and life-course data: The importance of duration», *Sociological Methods and Research*, vol. 38, n° 3, p. 365-388.
- HOLLISTER Matissa, 2009, «Is optimal matching suboptimal?», *Sociological Methods and Research*, vol. 38, n° 2, p. 235-264.
- KREUTER Frauke et KOHLER Ulrich, 2009, «Analyzing contact sequences in call record data: Potential and limitations of sequence indicators for nonresponse adjustments in the european social survey», *Journal of Official Statistics*, vol. 25, n° 2, p. 203-226.
- KUHN Thomas, 1962, *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press. Trad. fr.: *La structure des révolutions scientifiques*, trad. par Laure Meyer, Paris, Flammarion, 2012.

- LAKATOS Imre et MUSGRAVE Alan (eds.), 1970, *Criticism and the Growth of Knowledge*, Cambridge, Cambridge University Press.
- LEMERCIER Claire, 2005, « Les carrières des membres des institutions consulaires parisiennes au XIX^e siècle », *Histoire et Mesure*, vol. 20, n^{os} 1-2, p. 59-95.
- LESNARD Laurent, 2010, « Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns », *Sociological Methods and Research*, vol. 38, n^o 3, p. 389-419.
- 2008, « Off-scheduling within dual-earner couples: An unequal and negative externality for family time », *American Journal of Sociology*, vol. 114, n^o 2, p. 447-490.
- LESNARD Laurent et SAINT POL Thibaut de, 2006, « Introduction aux méthodes d'appariement optimal (*optimal matching analysis*) », *Bulletin de méthodologie sociologique*, vol. 90, p. 5-25.
- LEVINE Joel, 2000, « But what have you done for us lately? Commentary on Abbott and Tsay », *Sociological Methods and Research*, vol. 29, n^o 1, p. 34-40.
- MASCLET Camille, en cours, *La transmission familiale du féminisme. Le cas des militantes du Mouvement de libération des femmes et de leurs enfants*, thèse de doctorat sous la dir. d'Olivier Fillieule, Université de Lausanne.
- PICCARRETA Raffaella et LIOR Orna, 2010, « Exploring sequences: A graphical tool based on multi-dimensional scaling », *Journal of the Royal Statistical Society: Series A*, vol. 173, n^o 1, p. 165-184.
- POLLIER Alexandre et JOYE Dominique, 2014, « Patterns of contact attempts in surveys », dans Philippe Blanchard, Felix Bühlmann et Jacques-Antoine Gauthier (eds.), *Advances in Sequence Analysis: Theory, Method, Applications*, Dordrecht, Springer, p. 285-304.
- ROBETTE Nicolas, 2010, « The diversity of pathways to adulthood in France: Evidence from a holistic approach », *Advances in Life Course Research*, vol. 15, p. 89-96.
- ROBETTE Nicolas et BRY Xavier, 2012, « Harpoon or bait? A comparison of various metrics in fishing for sequence patterns », *Bulletin de méthodologie sociologique*, vol. 116, n^o 1, p. 5-24.
- ROBETTE Nicolas et THIBAUT Nicolas, 2008, « L'analyse exploratoire de trajectoires professionnelles. Analyse harmonique qualitative ou appariement optimal? », *Population*, vol. 64, n^o 3, p. 621-646.
- ROHWER Goetz et POETTER Ulrich, 2007, *Transition Data Analysis* [logiciel], Ruhr University, Bochum, en ligne : www.stat.ruhr-uni-bochum.de/tda.html (consulté en mars 2016).
- SCHLICH Robert, 2003, « Homogenous groups of travellers », 10th International Conference on Travel Behaviour Research, 10-15 août, Lucerne.

- STOVEL Katherine, 2001, « Local sequential patterns : The structure of lynching in the Deep South, 1882-1930 », *Social Forces*, vol. 79, n° 3, p. 843-880.
- STUDER Matthias, RITSCHARD Gilbert, GABADINHO Alexis et MÜLLER Nicolas S., 2011, « Discrepancy analysis of state sequences », *Sociological Methods and Research*, vol. 40, n° 3, p. 471-510.
- WIDMER Éric, LEVY René, POLLIEN Alexandre, HAMMER Raphaël et GAUTHIER Jacques-Antoine, 2003, « Entre standardisation, individualisation et sexualisation. Une analyse des trajectoires personnelles en Suisse », *Revue suisse de sociologie*, vol. 29, n° 1, p. 35-67.
- WILSON Clarke, 1998, « Analysis of travel behaviour using sequence alignment methods », *Transportation Research Record*, vol. 1645, p. 52-59.
- WILSON Matthew, 2014, « Governance built step-by-step : Analyzing sequences to explain democratization », dans Philippe Blanchard, Felix Bühlmann et Jacques-Antoine Gauthier (eds.), *Advances in Sequence Analysis : Theory, Method, Applications*, Dordrecht, Springer, p. 213-227.
- WU Lawrence L., 2000, « Some comment on "Sequence analysis and optimal matching methods in sociology: Review and prospect" », *Sociological Methods and Research*, vol. 29, n° 1, p. 41-64.

