

Du prosélytisme à la sécularisation

Le processus de diffusion de l'« Optimal Matching Analysis »

Nicolas Robette

Printemps (CNRS-UVSQ, UMR 8085)

Version du 22 mai 2012

La recherche en sciences sociales est liée depuis ses origines aux développements des mathématiques et des méthodes statistiques (Martin, 2002). Cette relation n'est pas univoque : certaines innovations statistiques ont été réalisées en réponse à des problèmes posés par les sciences sociales, par des chercheurs de l'un ou l'autre domaine (Clogg, 1992) ; mais le phénomène majeur reste l'importance des emprunts méthodologiques des sciences sociales aux statistiques. On pense par exemple aux modèles de durée, importés de la biologie et des sciences industrielles par Nancy Tuma, ou à l'analyse de réseaux sociaux par *blockmodeling* issue de la physique et appliquée à la sociologie par Harrison White. Certaines méthodes intègrent rapidement la boîte à outils des sciences sociales, quand d'autres restent à l'état de propositions, ne faisant pas l'objet d'une large appropriation. C'est notamment le cas de l'utilisation des « chaînes de vacance » (*vacancy chains*), suggérée par le même Harrison White : les sciences sociales ne se saisissent pas nécessairement - ou pas également - de ces emprunts méthodologiques. Dans ce contexte, les méthodes d'appariement optimal (« *optimal matching analysis* », AO) offrent un exemple d'importation relativement « réussie ».

Proposées par Andrew Abbott dans les années 1980, elles sont alors couramment utilisées en bio-informatique pour l'analyse des molécules d'ADN. Pendant plus d'une dizaine d'années, elles restent principalement liées aux travaux d'Abbott. Mais dès le début des années 2000, on observe une inflexion dans la diffusion de la méthode : elle est de plus en plus fréquemment employée et s'autonomise des travaux d'Abbott. C'est cette diffusion que nous tenterons d'éclairer ici, en l'envisageant comme un processus. La « carrière » de l'OM dans le domaine des sciences sociales se présente en effet comme la rencontre de diverses trajectoires, inscrites dans différentes échelles spatiales et temporelles, qui se croisent et parfois s'entrelacent : le parcours biographique et professionnel d'Abbott, l'évolution du champ des sciences sociales, de ses structures sociales et cognitives (Heilbron, 2003), et en particulier l'espace des méthodes¹.

¹ Ce travail est basé sur une revue de littérature exhaustive (ou presque) des publications sur l'appariement optimal, sur les travaux d'Abbott, mais aussi sur ma propre pratique de la méthode et de la réception de ses résultats, la participation à divers séminaires et colloques spécialisés et plus largement « l'observation participante » dans différents laboratoires de démographie et de sociologie. Je remercie Pernelle Issenhuth et Christophe Imbert et les éditeurs de cet ouvrage pour leurs relectures avisées.

L'appariement optimal dans le parcours d'Abbott

A l'orée des années 1980, Andrew Abbott réalise à l'université de Chicago une thèse de sociologie historique sur l'hôpital psychiatrique et ses professions, sous la direction de Morris Janowitz. En 1981, il donne un cours de sociologie historique aux étudiants de troisième cycle à l'université Rutgers (New Jersey). Il décide de thématiser ce cours et se concentre finalement sur les questions de séquence et d'ordre (Abbott, 2011b, p. 4). Ces réflexions sur les temporalités et les séquences vont nourrir bon nombre de ses travaux pendant une vingtaine d'années. Il publie ainsi des recherches qui mêlent réflexions théoriques et méthodologiques, notamment sur la conceptualisation des séquences et des événements qui les composent (ou « colligation »), mais aussi sur les moyens de les mesurer. Par l'intermédiaire d'un ancien camarade d'études, Abbott se met en contact avec Joseph Kruskal, statisticien renommé aux laboratoires Bell, et qui développe alors un ensemble de techniques pour l'analyse de séquences : les méthodes d'appariement optimal (Abbott, 2001b, p. 9). En 1986, Abbott publie avec l'anthropologue John Forrest le premier article de sciences sociales utilisant l'AO, qui analyse l'évolution des traditions locales en Angleterre, à partir des séquences de pas de danses traditionnelles.

Couramment utilisé en bio-informatique pour l'analyse des molécules d'ADN, l'appariement optimal est un algorithme de mesure de la similarité, *i.e.* du degré de ressemblance entre des séquences. La comparaison de l'ensemble des séquences d'un corpus est suivie le plus souvent par la classification de ces séquences : les séquences qui se ressemblent le plus sont regroupées, aboutissant finalement à la construction d'une typologie. Autrement dit, il s'agit d'explorer des corpus de données séquentielles et d'y identifier des régularités. Pour mesurer la similarité entre deux séquences, on transforme l'une des séquences en l'autre, en insérant ou supprimant des éléments à l'intérieur de la séquence, ou en substituant des éléments à d'autres. La dissimilarité entre les deux séquences est finalement égale au nombre minimal d'opérations (insertion, suppression, substitution) nécessaires à la transformation d'une séquence en l'autre. Il est de plus possible de pondérer les différents types d'opérations, selon que l'on souhaite privilégier la contemporanéité des séquences ou l'ordonnancement des éléments qui les composent².

L'usage que fait Abbott de la méthode est à replacer dans une réflexion plus large sur ce qu'il appelle la « réalité linéaire générale » (*general linear reality*) (Abbott, 2001b). Selon lui, le « cadre méthodologique » (*methodological manifold*) des sciences sociales est structuré par un ensemble d'oppositions, entre quantitatif et qualitatif, positivisme et interprétation, etc. (Abbott, 2001a, p. 28). Ces oppositions entretiennent des « affinités électives », la plus profondément ancrée associant d'une part le positivisme à l'analyse, et d'autre part le récit à l'interprétation. Abbott souhaite briser ces affinités en réintroduisant une dimension narrative dans le positivisme³. Il s'agit de proposer une alternative au paradigme des variables, qui domine l'empirisme quantitatif, et à ses présupposés implicites (Fabiani, 2003). L'analyse de séquences fournit un ensemble d'outils permettant de développer cette alternative, parmi lesquels Abbott va privilégier l'AO.

Durant les années 1980 et 1990, Abbott publie - seul, avec des collègues ou des étudiants - des

² Pour une présentation plus détaillée de la méthode, voir par exemple MacIndoe & Abbott, 2004 ou Robette, 2011.

³ Il nomme cette rupture le « positivisme narratif » (Abbott, 1992), avant d'abandonner cette terminologie.

articles théoriques sur l'approche par les « configurations narratives », des recensions et discussions sur les méthodes d'analyse de séquences, et diverses applications de l'AO à des objets et dans des disciplines variés. On peut voir là la première période de la diffusion de l'AO, celle du « prosélytisme » : le processus repose entièrement sur Abbott, les rares applications qu'il ne signe (ou ne co-signe) pas étant presque toutes le fait de jeunes chercheurs qu'il a suivis durant leur thèse⁴.

Bien qu'hétérodoxe, l'innovation proposée par Abbott rencontre un certain écho, sans doute lié à la position que le sociologue occupe dans le champ académique. En effet, lorsqu'Abbott introduit l'AO durant les années 1980, il possède déjà un capital d'autorité scientifique important, selon les termes de la sociologie des sciences de Pierre Bourdieu (Bourdieu, 2001, p. 113) : c'est un chercheur reconnu, particulièrement pour sa sociologie des professions. Dans le prolongement de sa thèse, il publie de nombreux articles dans des revues prestigieuses, comme l'*American Journal of Sociology*, *Social Forces* ou *Work & Occupations*, et son ouvrage *The system of professions*, paru en 1988, est reconnu rapidement comme majeur. Il possède aussi un certain capital de pouvoir sur le monde scientifique (*ibid.*, p. 113), une position institutionnelle favorable à la diffusion de l'AO. Abbott est professeur associé à l'université Rutgers à partir de 1986, puis professeur titulaire à l'université de Chicago à partir de 1991. Il peut bénéficier de la collaboration de doctorants et d'assistants de recherche, et siège dans les comités de rédaction de plusieurs revues qui publieront des recherches utilisant l'AO (*American Journal of Sociology*, *Sociological Methods & Research*, *Work & Occupations*).

Mais la position d'Abbott ne se résume pas à son capital scientifique : il possède aussi des dispositions à la subversion des règles du jeu académique. A la manière des héritiers (Bourdieu & Passeron, 1964, p. 27), il se montre à la fois investi et désinvolte à l'égard du travail de recherche, rapport sans doute autorisé et produit par ses origines sociales et son parcours scolaire. A l'université, il se considère comme un intellectuel en puissance (*would-be intellectual*) et choisit la sociologie parce qu'il pense que cela l'autorisera à étudier ce qu'il souhaite, comme il le souhaite (Abbott, 2001b, p. 2). Il revendique un certain éclectisme, voire des « troubles déficitaires de l'attention intellectuelle » (*ibid.*, p. 2). Il souligne régulièrement l'importance du plaisir, du « fun » (Abbott, 2004), qui passe par la liberté de ton et l'excitation à réaliser des « coups méthodologiques » (Abbott, 2001a, p.231), dont l'analyse de séquences est un exemple.

Toutefois, les difficultés rencontrées par Abbott pour publier sa recherche sur les carrières de musiciens allemands de l'âge baroque (Abbott, 2001b, p.13) laisse penser que sa position dans le champ scientifique n'était pas encore un atout suffisant pour s'imposer à la fin des années 1980, et surtout qu'il n'est pas aisé de s'attaquer au paradigme dominant, qui plus est lorsqu'à l'hétérodoxie de l'approche est associée la « futilité » de l'objet⁵.

⁴ Durant ces années, Abbott co-publie notamment avec Alexandra Hrycak, Emily Barman et Heather MacIndoe, dont il dirige les thèses. Il dirige aussi la thèse de Mary Blair-Loy et fait partie des jurys de Katherine Stovel et Daniel Carpenter. Au final, parmi les articles recensés, deux seulement ne semblent pas avoir été associés à des interactions directes avec Abbott : ils ont été écrits par Tak Wing Chan et Brendan Halpin.

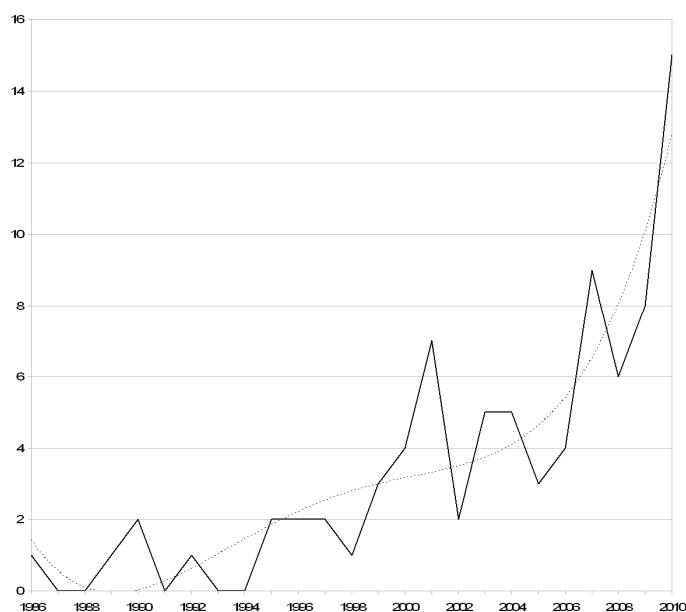
⁵ James Coleman déclare ainsi à Abbott : « Personne ne prêtera attention à ce que tu dis tant que tu écriras sur des musiciens allemands morts » (« *Nobody's gonna pay any attention to what you're saying as long as you write about dead German musicians* ») (Abbott, 2001b, p. 13).

Du prosélytisme à la sécularisation

Aux environs de l'année 2000 débute ce que l'on peut considérer comme une deuxième phase dans le processus de diffusion de l'appariement optimal. Le nombre de publications est croissant, voire exponentiel (voir figure 1), et on observe une autonomisation de la diffusion par rapport à Abbott : le cercle de chercheurs utilisant la méthode s'élargit, indépendamment du sociologue de Chicago. C'est la « sécularisation » de l'AO.

Ce premier constat fait, quelles sont les modalités de la diffusion de l'appariement optimal ? Tout d'abord, on assiste selon certains auteurs à un tarissement du paradigme des variables dans la sociologie quantitative anglo-saxonne (Abbott, 2001b ; Savage, 2009). Si les modèles stochastiques ont certes permis pendant un temps d'enrichir significativement les connaissances sur divers objets sociologiques, il semble que ces méthodes soient arrivées au bout de leur potentiel heuristique : des « anomalies » apparaissent, *i.e.* un certain nombre d'énigmes ne sont plus résolues - ou imparfaitement - à l'aide des approches majoritaires (Kuhn, 1962). On se trouve donc dans une configuration propice à l'apparition d'une nouvelle approche méthodologique.

Figure 1 – Nombre de publications de l'appariement optimal entre 1986 et 2010



Lecture : en trait plein figure le nombre d'articles publiés chaque année ; en pointillés une courbe de tendance (polynômiale).

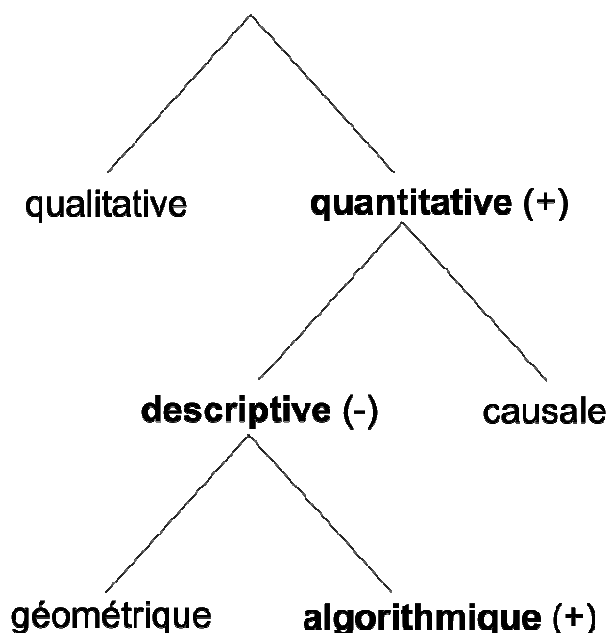
Cette configuration est concomitante du développement d'un « agenda », de nouvelles énigmes à résoudre (*ibid.*), cette fois non pas sur le plan des méthodes mais sur celui des objets de recherche et des données empiriques existantes. C'est particulièrement le cas dans l'étude des carrières professionnelles et dans celle des parcours de vie dans la recherche anglo-saxonne. Durant les années 1980, ces deux domaines bénéficient conjointement de la disponibilité croissante d'importants corpus de données et de l'introduction de l'*event history analysis* (EHA), ensemble de techniques permettant de modéliser statistiquement des durées ou des probabilités d'occurrence d'événements (voir par exemple Tuma *et al*, 1979). L'EHA

est en quelque sorte une extension du paradigme des variables à des données diachroniques : elle s'inscrit donc naturellement dans le pôle *mainstream* de l'empirisme quantitatif et se diffuse très rapidement. Or c'est précisément sur ces thématiques que la majeure partie des travaux utilisant l'AO va être publiée : sur les 83 articles recensés, 32 étudient des carrières professionnelles et 22 des parcours de vie (soit près des deux tiers au total). A partir de ces deux spécialités, l'usage de l'AO se répand de proche en proche, par capillarité, à des sous-champs thématiques voisins (par exemple les trajectoires financières ou résidentielles), à des disciplines voisines (comme la démographie ou les sciences de l'éducation) ou à d'autres aires géographiques (notamment l'Allemagne ou la Suisse)⁶.

L'idée d'agenda peut aussi être pensée en termes de données. En effet, depuis 1945, la complexité des données quantitatives disponibles en sciences sociales n'a cessé de croître. Adrian Raftery propose donc de distinguer trois générations de méthodes statistiques en sociologie, selon le type de données qu'elles analysent (Raftery, 2001). La première génération, qui commence dans les années 1940, est centrée sur les tableaux de contingence, avec les mesures d'association et les modèles log-linéaires. La seconde génération, apparue dans les années 1960, analyse les données d'enquêtes individuelles et regroupe des modèles de régression (LISREL, EHA). Enfin, la troisième génération, qui émerge à la fin des années 1980, traite de données « atypiques », par exemple des textes, des données relationnelles, spatiales ou séquentielles. Les données sur les trajectoires biographiques, notamment, se sont particulièrement développées avec l'essor des enquêtes longitudinales dites de « panel », interrogeant à intervalles répétés les mêmes individus ou ménages, souvent plusieurs milliers (par exemple le *Panel Study on Income Dynamics*, créé en 1968 aux Etats-Unis, ou le *British Household Panel Survey*, mis en place en 1991 au Royaume-Uni). Or le traitement de ces données séquentielles est précisément ce que permet l'AO : sa diffusion serait donc en partie liée à l'existence de nouvelles formes de données.

⁶ Cette capillarité est d'autant plus efficace que les sous-champs thématiques, disciplines ou aires géographiques voisins sont suffisamment autonomes pour que l'introduction de l'AO puisse y prendre les traits de la nouveauté et offrir ainsi une plus-value symbolique non négligeable.

Figure 2 – Représentation fractale de l'espace des méthodes



On l'a vu, l'AO occupe donc une position particulière dans l'espace des méthodes en sciences sociales. Cet espace peut être représenté sous forme de fractales (figure 2). Le modèle des fractales développé par Abbott pour rendre compte de l'évolution et de la différenciation des disciplines académiques, notamment, est fondé sur le fait que les dichotomies – les oppositions binaires sociales ou culturelles – se répètent à l'identique à l'intérieur d'elles-mêmes, comme les fractales géométriques (2001a). Appliqué à l'espace des méthodes (*ibid.*, p. 11), on observe tout d'abord, au premier niveau de la fractale, l'opposition entre méthodes « quantitatives » et « qualitatives »⁷ : l'AO se situe dans la branche dominante, au moins en ce qui concerne les sciences sociales anglo-saxonnes⁸. Le second niveau correspond à la différenciation fractale entre le paradigme des variables, autrement dit la modélisation causale (régressions, etc.), et une approche par les configurations, en quelque sorte plus « qualitative », qui vise avant tout à explorer et décrire⁹ : cette fois, l'AO se trouve dans le pôle dominé. Mais cette position peut avoir un effet favorable à la diffusion de la méthode, dans le sens où son utilisation, dans ce contexte, peut être un investissement « payant », un moyen pour les nouveaux entrants dans le champ scientifique de se distinguer et de subvertir les règles du jeu et les systèmes de classement. La hiérarchie sociale des objets de recherche décrite par Bourdieu (1975)¹⁰ semble ici transposable aux méthodes. Ces nouveaux entrants sont les agents de ce qu'Abbott appelle un « paradigme générationnel » (*ibid.*, p.23), c'est-à-

⁷ Il n'est pas question ici de discuter de la pertinence de ces oppositions mais simplement de prendre acte de leur caractère structurant dans le champ scientifique.

⁸ La situation est différente en France, où les approches quantitatives sont plus marginales.

⁹ On retrouve des oppositions analogues chez Desrosières (2001) – réalisme métrologique vs conventions d'équivalence – ou Cibois (1981) – méthodes durkheimiennes vs wéberiennes.

¹⁰ « [...] le champ des objets de recherche possibles tend toujours à s'organiser selon deux dimensions indépendantes, c'est-à-dire selon le degré de légitimité et selon le degré de prestige à l'intérieur des limites de la définition légitime » (Bourdieu, 1975, p. 5).

dire le développement d'un paradigme résultant de l'opposition de jeunes chercheurs à la génération antérieure et dominante. Un rapide examen bibliométrique semble confirmer cette proposition : parmi les auteurs d'un article utilisant l'AO, plus de la moitié était en thèse ou avaient soutenu depuis moins de 5 ans lors de la publication de l'article ; plus des trois quarts parmi les auteurs ayant publié un article seul (22 sur 27). Ce sont bien les nouveaux entrants dans le champ des sciences sociales qui forment la majeure partie des utilisateurs de l'AO. Pour finir, le dernier niveau de la fractale oppose deux « familles » de méthodes d'analyse de séquences, et sera développé plus loin.

A la différence de l'analyse des réseaux sociaux, par exemple, qui possède une association et une revue spécifiques (Freeman, 2004), le groupe social formé par les utilisateurs de l'AO n'est pas adossé à des modes d'organisation très formalisés. Toutefois, des liens se créent et une structure sociale émerge progressivement. Ces liens intellectuels sont avant tout indirects et caractérisés par « l'influence exercée par le travail de certains chercheurs sur celui des autres » (Crane, 1969). En ce sens, le travail d'Abbott constitue une référence partagée par l'ensemble des utilisateurs de l'AO : tous les articles mentionnant l'AO citent au moins un article d'Abbott. La circulation de l'influence opère principalement à travers les publications dans des revues scientifiques, dont certaines – que l'on pourrait qualifier d'« optimal-matching friendly » - exercent un rôle central. C'est par exemple le cas de l'*American Journal of Sociology* à partir de 1990 et la publication de l'article d'Abbott et Hrycak sur les musiciens allemands (6 articles), puis de la *European Sociological Review* dès 1998 (4 articles), de *Sociological Methods & Research* à partir de 2000 (13 articles et 2 numéros spéciaux), du *Journal of the Royal Statistical Association Series A (Statistics in Society)* à partir de 2002 (5 articles) et d'*Advances in Life Course Research* à partir de 2009 (4 articles). Ces cinq revues centrales regroupent plus du tiers des articles utilisant l'AO, les autres articles étant répartis dans une quarantaine d'autres revues plus périphériques. Cette « dispersion bibliographique » (*reference scattering*) entre un centre et une périphérie est caractéristique d'un domaine de recherche à la fois relativement ouvert et favorable à une certaine cumulativité (Crane, 1969). De même que la dispersion bibliographique, la productivité des utilisateurs de l'AO n'est pas uniforme : parmi les 120 auteurs d'un article avec de l'AO, 12 ont écrit trois articles ou plus, ce qui représente plus du tiers de la production totale. La minorité d'auteurs les plus productifs exerce un « leadership informel » sur le groupe social des utilisateurs de l'AO (*ibid.*). Au final, ceux-ci forment donc un « cercle social » (Kadushin, 1966 ; Crane, 1969)¹¹ unique et complet, dans le sens où il comprend l'ensemble des utilisateurs.

Cependant, les contacts se font aussi plus directs, en particulier durant la deuxième moitié des années 2000. Des séminaires fermés sont organisés (en l'absence d'Abbott) : entre 2007 et 2008, l'ensemble des chercheurs français ayant publié des articles avec de l'AO font ainsi partie d'un groupe de travail, sous l'égide du Réseau Thématique « Méthodes » (RT20) de l'Association Française de Sociologie ; en septembre 2007, Brzinsky-Fay et Kohler réunissent une vingtaine de chercheurs à l'université de Berlin lors du workshop *Sequence Analysis in the Social Sciences* ; en juin 2012 se tient la *Lausanne Conference on Sequence Analysis*, organisée Blanchard, Buhlmann et Gauthier. L'AO s'introduit aussi dans des colloques ouverts : en 2008 et 2012, les 7ème et 8ème *International Conferences on Social Science*

¹¹ Selon Kadushin, un « cercle social » est un groupe d'individus caractérisé par des interactions denses, indirectes et basées sur un intérêt commun, une faible institutionnalisation et un leadership informel (Kadushin, 1966).

Methodology du RC33 de l'*International Sociological Association* consacrent plusieurs séances à l'analyse de séquences. La formation de nouveaux utilisateurs s'institutionnalise avec la multiplication de séminaires et écoles d'été, aux États-Unis, en Angleterre, en France, en Allemagne...¹² Ces différents lieux sont le cadre d'interactions directes et informelles et de la constitution de « collèges invisibles », c'est-à-dire de groupes réunissant une « élite » de chercheurs productifs et en interaction dans ce domaine de recherche (Price & Beaver, 1966).

Un autre élément intervient dans la diffusion de l'AO : la question des moyens techniques de sa mise en œuvre. Compte tenu de la quantité de calculs effectués par l'algorithme, recourir à l'informatique est indispensable : les possibilités d'analyse sont donc contraintes par les développements informatiques et logiciels. Compte tenu de la puissance des ordinateurs disponibles, les premières applications des méthodes d'AO en sciences sociales portent sur des échantillons relativement petits (quelques dizaines de séquences) - 27 séquences de pas dans l'analyse de danses traditionnelles en Angleterre (Abbott & Forrest, 1986), 37 carrières professionnelles à Hong-Kong (Chan, 1995) ou 279 carrières de musiciens allemands au XVIIIème siècle (Abbott & Hrycak, 1990), par exemple - alors même que des enquêtes longitudinales auprès de plusieurs milliers d'individus existent depuis plusieurs décennies. Par la suite, la taille des corpus analysables et analysés augmente progressivement : à l'heure actuelle, les capacités techniques des ordinateurs ne constituent plus vraiment une entrave pour la plupart des bases de données disponibles.

Les contraintes techniques dans la diffusion de l'AO ne concernent pas uniquement la puissance des ordinateurs. Encore faut-il que l'algorithme soit transformé en un logiciel, la plupart des chercheurs n'ayant pas les ressources (en termes de compétences informatiques et/ou de temps) pour réaliser eux-mêmes un programme implémentant l'AO. Cette « *commodification* » s'est réalisée graduellement. Pour ses premières applications, Abbott emprunte une suite de logiciels, les *Belding Program Series*, programmée au milieu des années 1970 par David Bradley à la Long Beach State University, et utilisée par des biologistes pour étudier le langage des moineaux. Au début des années 1990, Abbott développe son propre programme, Optimize. De manière générale, durant la première phase de la diffusion de l'AO, les chercheurs utilisent des logiciels spécifiques, qu'ils ont eux-mêmes développés ou issus de la bio-informatique. Ces logiciels sont généralement peu conviviaux et relativement limités quant à la gamme d'analyses statistiques disponibles.

En 1997 apparaît TDA (pour *Transition Data Analysis*) : développé par Götz Rohwer et Ulrich Pötter, à l'université de Bochum (Allemagne), ce logiciel de statistiques est originellement conçu pour l'analyse de données longitudinales, en particulier l'*event history analysis*, mais il possède également un module permettant d'analyser des séquences par appariement optimal (Rohwer & Pötter, 2005). De plus, il est gratuit, très bien documenté et suffisamment puissant pour traiter de grandes bases de données. Toutefois, il ne suffit pas pour une analyse statistique complète, dans la mesure où certaines techniques, comme l'analyse des correspondances ou les classifications, en sont absentes. Il est pourtant utilisé dans plus d'une vingtaine d'articles entre 2001 et 2010 : la trajectoire de ce logiciel et la seconde phase de la diffusion de l'AO semblent largement coïncider.

En 2008, une équipe de l'Institute for Demographic and Life Course Studies (IDEMO), à

¹² Alors que la formation était principalement informelle et liée aux interactions entre Abbott et ses étudiants pendant la phase de prosélytisme (voir note 4).

l'Université de Genève, propose le package TraMineR, exclusivement consacré à l'analyse de séquences et à leur représentation graphique (Gabadinho *et al*, 2011). Ce travail est financé par la Swiss National Foundation for Scientific Research, dans le cadre du projet « Mining Event Histories ». Développé comme un module additionnel pour le logiciel de statistiques généraliste et libre R, TraMineR bénéficie donc de toutes les autres fonctionnalités de ce logiciel, qui permettent de réaliser relativement aisément l'ensemble des traitements nécessaires à une analyse de séquences complète (du codage à la classification). Il est puissant, bien documenté et fréquemment mis à jour (de nouvelles fonctions et des algorithmes alternatifs à l'AO y sont régulièrement ajoutés). De plus, ses auteurs participent à de nombreux séminaires de formation et écoles d'été dans différents pays. Le succès de TraMineR est spectaculaire, favorisé par celui de R : la moitié des articles publiés en 2010 et mettant en œuvre l'AO y font appel, et deux tiers de ceux publiés en 2011. Son utilisation déborde d'ailleurs les frontières des seules sciences sociales : la liste des membres de sa *mailing list* comprend ainsi des chercheurs en mathématiques appliquées ou en biologie. Comme le montrait Clogg (1992), les méthodes statistiques empruntées par les sciences sociales à d'autres disciplines y sont parfois enrichies de nouveaux développements et peuvent profiter à leurs communautés d'origine.

Un turning point ?

L'appariement optimal n'est pas la seule méthode à permettre la comparaison de séquences. Il existe même de nombreuses méthodes similaires (Robette, 2011), que l'on peut classer en deux « familles » (Lesnard & de Saint Pol, 2009). La première est « géométrique » et tire parti de la tradition d'analyse des données « à la française », ou analyse géométrique des données (Rouanet & LeRoux, 2011), en particulier l'analyse des correspondances. La seconde, « algorithmique », est basée sur des algorithmes souvent issus des sciences dites « dures » (statistique, biologie et informatique, principalement). Les méthodes géométriques sont cependant très peu employées. Tout d'abord, on peut y voir le reflet du faible poids de l'analyse géométrique des données en général. Les explications envisageables sont de différents ordres : quasi-monopole anglo-saxon « de type colonial » sur la statistique (Rouanet, 2008), alors même que l'analyse des données s'est développée à partir des travaux français de Benzécri ; frontières linguistiques ; lacunes de la formation en géométrie aux États-Unis, qui entrave une vision « spatiale » des problèmes... En France, le succès de ces méthodes dans les sciences sociales durant les années 1970 et 1980, souvent associé aux travaux de Pierre Bourdieu et à la théorie des champs, a connu un certain reflux. En ce qui concerne plus spécifiquement l'analyse de séquences, l'analyse géométrique des données a été principalement employée en France dans des études sur l'insertion professionnelle menées dans le cadre des enquêtes du Céreq¹³, peu diffusées dans le champ scientifique. A l'étranger, les références sont rares (Van Heijden, 1987).

Les méthodes algorithmiques, dont fait partie l'AO, bénéficient quant à elles d'un écho nettement plus large, ce qui traduit notamment leur origine : les algorithmes, directement repris ou inspirés de la bio-statistique ou de l'informatique, peuvent apparaître plus solides, plus rigoureux, en d'autres termes plus « quantitatifs ». L'AO se trouve donc dans le pôle

¹³ Centre d'Études et de REcherches sur les Qualifications

dominant des méthodes d'analyse de séquences (voir figure 2).

A l'intérieur de ce pôle, diverses méthodes existent et l'AO est de très loin la plus utilisée. Les raisons peuvent être recherchées du côté de ses qualités intrinsèques : la possibilité de jouer sur les paramètres de l'algorithme (i.e. le poids des opérations de substitution, insertion et suppression) lui confère beaucoup de souplesse et d'adaptabilité, selon que l'on souhaite privilégier telle ou telle dimension de la temporalité du processus étudié (la contemporanéité ou l'ordre des événements). Mais cette souplesse est aussi parfois perçue comme une faiblesse, faisant de la méthode une boîte noire à la merci de l'arbitraire du chercheur. De plus, il n'existe pas de comparaison démontrant la supériorité de l'une ou l'autre des techniques d'analyse de séquences (Robette, 2012). La suprématie de l'AO a donc aussi à voir avec des mécanismes externes - intellectuels, culturels, institutionnels et sociaux -, qui participent d'un processus de légitimation (Ben-David & Collins, 1966 ; Lamont, 1987).

On a vu que l'on pouvait distinguer deux périodes dans le processus de diffusion des méthodes d'appariement optimal, et l'année 2000 fait figure de transition entre ces deux périodes¹⁴, de manière symbolique et en partie arbitraire. Toutefois, on peut aussi faire l'hypothèse que cette année constitue un véritable « turning point » (Abbott, 1997), dans le sens d'un changement de pente, d'une redistribution des probabilités (Passeron, 1989) de publication de l'AO.

En effet, la revue *Sociological Methods & Research* publie en 2000 un numéro spécial sur l'AO. Abbott et Angela Tsay, l'une de ses étudiantes, y proposent un large état des lieux sur l'utilisation de ces méthodes en sciences sociales, en abordant les questions de données, de codage, de temporalités, de paramétrage de l'algorithme et de stratégies d'analyse, et concluent de manière optimiste sur les promesses ouvertes par l'analyse de séquences. Suivent deux articles très critiques de Joel Levine (Levine, 2000) et Lawrence Wu (Wu, 2000), tous deux membres du comité éditorial de la revue¹⁵. Le dossier s'achève par une réponse d'Abbott à Levine et Wu (Abbott, 2000).

Levine se place fermement du point de vue du pôle *mainstream* de la fractale des méthodes, i.e. celui de la réalité linéaire générale, et reproche principalement à l'AO de ne pas répondre aux standards des modèles stochastiques. Wu, qui est spécialiste de l'*event history analysis*, adopte le même point de vue mais émet également des critiques plus ciblées, portant sur des aspects précis de la méthode, comme la signification sociologique des opérations de substitution, d'insertion et de suppression d'éléments à l'intérieur des séquences, ou la prise en compte de l'ordre des événements qui composent les séquences. Abbott répond point par point à l'ensemble de ces critiques. Il rectifie ce qu'il considère comme des malentendus sur le fonctionnement de l'AO, mais surtout il replace la méthode dans le cadre de l'opposition entre réalité linéaire générale et méthodes narrativo-descriptives : le fait d'évaluer l'AO à l'aune des fondements des méthodes *mainstream* invalide *de facto* ces critiques. Abbott conclut d'ailleurs sa réponse par un éloquent « *I'm bloody but unbowed* »¹⁶.

Au-delà du contenu de ces débats, leur existence même semble produire un effet sur la diffusion de la méthode. Tout se passe en effet comme si la discussion dans le numéro spécial

¹⁴ Voir aussi la contribution de Philippe Blanchard dans ce volume.

¹⁵ Abbott deviendra à son tour membre du comité éditorial en 2007.

¹⁶ « Je suis ensanglanté mais pas vaincu ».

de *Sociological Methods & Research* marquait la légitimation de l'AO comme méthode « sérieuse », rigoureuse, digne de figurer parmi les outils du sociologue quantitatif. Cette hypothèse est appuyée par l'observation des citations des articles critiques de Levine et Wu : leur fréquence évolue de manière parallèle à celui des publications utilisant l'AO (bien qu'à un niveau moins élevé). En revanche, ce n'est pas le cas de la réponse d'Abbott aux critiques¹⁷. Il semble donc que ce ne soit pas le fond de la discussion qui importe, mais son existence même. Citer la discussion de la méthode permet paradoxalement de souligner son appartenance au corpus des méthodes légitimes et sa scientificité, indépendamment des points précis sur lesquels portent les critiques.

Enfin, la légitimité conférée à l'AO par le capital symbolique d'Abbott et lors de ce *turning point*, ainsi que sa position dans l'espace des méthodes, se combinent et se renforcent mutuellement. Sa diffusion va croissante, au détriment d'autres alternatives similaires, dans une dynamique d'auto-renforcement, une sorte d'« effet saint Matthieu » (Merton, 1968).

Différentes formes d'appropriation de l'appariement optimal

Les travaux de sciences sociales utilisant l'appariement optimal ne s'en sont pas saisis uniformément. La place de la méthode varie, du simple outil au service de la réponse à une question de recherche, à l'objet central de la discussion. L'AO peut en outre être mobilisé de manière classique, pour construire une typologie de séquences selon l'exemple de l'article d'Abbott sur les musiciens allemands par exemple, ou de manière plus innovante, en apportant des modifications techniques à la méthode ou pour produire des résultats ne prenant pas la forme de typologies. A grands traits, on peut distinguer quatre formes idéal-typiques d'appropriation de l'AO.

L'usage « innovateur » renouvelle les objets étudiés ou à la manière de mettre en œuvre la méthode. Il s'agit par exemple d'analyser des données d'emplois du temps, là où la majorité des applications traitent de carrières professionnelles ou de parcours de vie (Lesnard, 2010). Cela peut encore consister à mesurer et expliquer l'écart à une séquence « normale », alors que plus de 90% des articles s'attachent à identifier les principaux types de séquences d'un corpus : quels sont par exemple les facteurs qui déterminent l'éloignement à une carrière complète d'emploi à temps plein (Scherer, 2001) ? Ces usages « innovateurs » ouvrent de nouvelles perspectives, et peuvent par la suite faire fonction d'exemples, de matrices dont s'inspireraient des continuateurs.

Assez proche de l'« innovateur », l'usage « techniciste » concentre l'innovation sur les aspects les plus techniques de l'AO : c'est le fonctionnement de la méthode qui est perfectionné, amendé ou complété. C'est notamment le cas des travaux qui proposent des modifications des opérations de substitution, d'insertion ou de suppression (par exemple Halpin, 2010).

L'usage « stratégique » a une vocation instrumentale : il s'agit de capitaliser le prestige associé à la relative nouveauté de la méthode, en se positionnant en marge du tout venant des approches *mainstream*. Il serait difficile (et peu amène) de citer des exemples, tant il est vrai que la recherche de distinction traverse l'ensemble de la production scientifique. Ici, le relatif

¹⁷ D'après *Google Scholar*, les articles de Levine et Wu sont cités respectivement 22 et 52 fois dans des revues à comité de lecture ; la réponse d'Abbott ne compte que 3 citations.

succès de l'AO lui confère cette fonction, ce qui peut paradoxalement conduire à une sorte de nouveau conformisme.

Enfin, l'usage « pragmatique » a un caractère utilitaire. Une fois pris acte de l'existence de ce nouvel outil, l'AO est mis en œuvre lorsqu'il est susceptible d'apporter un éclairage intéressant à l'objet de recherche. Ce mode d'appropriation est également présent dans la plupart des travaux. En effet, les différents usages de la méthode ne sont pas exclusifs. Au contraire, ils se combinent le plus souvent dans un même compte-rendu de recherche. Les travaux de Laurent Lesnard (2010) en fournissent de nouveau un exemple. Ce sociologue a doublement innové en appliquant les méthodes d'analyse de séquences à des données d'emploi du temps et en proposant une version modifiée de l'algorithme d'AO, le *dynamic Hamming matching*, ce qui a permis d'enrichir la connaissance des emplois du temps des Français. Son capital symbolique dans le champ scientifique s'en est trouvé renforcé, ses recherches faisant référence tant sur le plan de l'objet étudié que sur le plan méthodologique.

Par ailleurs, les différentes formes d'appropriation de l'AO s'inscrivent dans le processus de diffusion de la méthode et leur importance relative varie au cours de ce processus. Une plus grande profondeur historique serait nécessaire à l'objectivation de ces fluctuations mais on peut tout de même émettre quelques hypothèses. Par exemple, il est probable que les usages « innovateurs » et « technicistes » soient surtout présents au début de la diffusion : d'une part, le mode d'utilisation d'une méthode a tendance à se stabiliser, à se cristalliser autour d'un ou plusieurs travaux exemplaires ; d'autre part, les innovations potentielles ne sont pas infinies et peuvent à terme s'épuiser. L'usage « stratégique » est principalement efficace durant la phase intermédiaire de la diffusion, lorsque la méthode déborde le cercle des pionniers tout en n'étant pas largement connue et reconnue. La plus-value symbolique de l'AO devrait s'émousser à mesure que la méthode s'impose comme outil légitime et standard de l'analyse quantitative. On peut enfin penser que l'usage « pragmatique » se diffuse parallèlement à la méthode, tendant à devenir plus ordinaire au fil de la routinisation de l'appropriation de l'AO et du vieillissement social des pionniers et des nouveaux utilisateurs..

Après avoir essayé de mettre au jour les différentes dimensions du processus de diffusion de l'appariement optimal, on peut se demander dans quelle mesure ces dimensions forment une grille d'analyse généralisable. Répondre à cette question nécessiterait un travail approfondi de comparaison avec diverses autres méthodes et l'histoire de leur importation puis de leur éventuelle adoption dans les sciences sociales. L'étude de quelques exemples (*event history analysis, qualitative comparative analysis*) ne peut être développée dans le cadre de ce chapitre mais elle permet d'ébaucher une hypothèse : la grille d'analyse proposée s'applique plus particulièrement aux approches situées dans une position intermédiaire dans l'espace des méthodes, hors du pôle dominant de la réalité linéaire générale mais suffisamment « positivistes » pour ne pas en être trop éloignées.

1986-2010 : bilan provisoire

En 2000, en conclusion de sa réponse aux critiques de Levine et Wu, Abbott, de manière provocatrice, donnait rendez-vous aux sceptiques 35 ans plus tard pour constater le succès de l'appariement optimal. Dix années plus tard, on est tenté de lui donner raison : la méthode est

en passe de s'imposer comme un outil d'analyse quantitative légitime en sciences sociales. On peut toutefois nuancer ce constat optimiste.

Il faut tout d'abord se garder de la tentation téléologique : il ne va pas de soi que le sens de la pente de la diffusion de l'AO reste inchangé dans les années à venir. Les rapports de force à l'intérieur de l'espace des méthodes peuvent évoluer dans une direction qui placerait l'AO dans une position moins favorable, annihilant ainsi la plus-value symbolique de son usage « stratégique ». L'exemple de l'analyse des correspondances illustre le fait que l'utilisation d'une méthode peut fluctuer avec les structures sociales et cognitives du champ scientifique, notamment les rapports de force et/ou de collaboration à l'intérieur et entre les disciplines (van Meter *et al*, 1994 ; Rosental, 1997 ; Desrosières, 2008)¹⁸. Mais, à tout le moins, un retour à la case départ ne semble guère possible : l'existence de l'AO est maintenant « durcie », incarnée dans un corpus de publications et des logiciels qui ne s'évanouiront pas quelle que soit l'appropriation effective de la méthode (ou son absence).

Ensuite, Abbott - toujours en 2000 - remarquait avec Levine qu'il n'existait pas d'application véritablement marquante (« *a blockbuster application* », p. 75), qui ferait significativement « bouger les lignes » sur une thématique donnée ; 12 ans plus tard, la situation n'a pas beaucoup évolué. Beaucoup des analyses de séquences menées durant ces années sont de grande qualité, mais peu marqueront une étape décisive dans la connaissance et l'appréhension de leurs objets d'étude. Mais c'est peut-être trop demander à une méthode dont l'ambition consiste simplement à identifier des régularités (*pattern searching*) : l'AO ne marque pas une révolution, un changement profond de paradigme (Kuhn, 1962). Abbott concède également que son éclectisme l'a éloigné des recherches et du prosélytisme qui auraient pu porter plus loin la méthode (Abbott, 2000, pp. 66-67).

A la rareté des applications « blockbuster » s'ajoute l'étroitesse du spectre des objets étudiés. Sur 77 articles exploitant des données empiriques, seuls 8 n'analysent pas des trajectoires individuelles (carrières, parcours de vie, etc.) ; parmi ceux-ci, 4 ont été écrits (ou co-écrits) par Abbott¹⁹. L'AO s'est finalement peu diffusé au delà de l'étude des carrières et des parcours de vie. L'agenda identifié par Abbott (2001b, p.61)²⁰ n'a pas rencontré un large public. Le « positivisme instrumental »²¹ dominant (Bryant, 1989 ; Ollion, 2011) inhibe peut-être l'imagination sociologique, empêchant de concevoir l'ensemble des applications potentielles de l'analyse de séquences (Mills, 1959 ; Abbott, 2001b, p. 63).

Plus généralement, alors que l'AO comme proposition méthodologique s'inscrivait chez Abbott dans un projet théorique plus large, ce cadre général de recherche n'est que très peu repris (les articles programmatiques d'Abbott sont d'ailleurs beaucoup moins cités que ses

¹⁸ L'analyse des correspondances aurait connu un âge d'or dans son association avec l'approche sociologique de Pierre Bourdieu, avant de pâtir des attaques contre les approches structurales.

¹⁹ Les articles signés par Abbott portent sur des pas de danses folkloriques, l'adoption de programmes de sécurité sociale par des états et la structure rhétorique d'articles scientifiques. Les autres analysent des trajectoires d'entreprises, des contacts avec des enquêtés lors d'une enquête par questionnaire, l'évolution historique des lynchages au Sud des États-Unis et la structure de manuels de physique et de sociologie.

²⁰ C'est-à-dire les situations où l'on étudie le déroulement d'un processus, avec peu de cas, une information collectée large et hétérogène ; les organisations, les professions, les révolutions ou les politiques par exemple.

²¹ Selon Bryant, « le "positivisme instrumental" [...] est "instrumental" dans la mesure où ce sont les instruments de recherche disponibles qui délimitent l'objet de la recherche, et "positiviste" en ce que cette autocontrainte des sociologues traduit leur volonté de se soumettre à une rigueur d'analyse comparable à celle qu'ils attribuent aux sciences naturelles » (Bryant, 1989).

applications empiriques). Comme dans le cas du paradigme des variables avant elle (Fabiani, 2003), la méthode s'autonomise de la théorie. Cette « sécularisation »²² illustre une fois encore la puissance hégémonique de la réalité linéaire générale.

Enfin, en ce qui concerne l'émergence du cycle fractal de différenciation entre *event history analysis* et AO, les deux branches semblent pour le moment coexister, avec une certaine porosité entre les deux approches. Mais cette porosité n'est pas symétrique : la branche dominante, celle de l'EHA, va intégrer, « ingérer » (Abbott, 2001a, pp. 18-20) certaines questions ou critiques soulevées par l'analyse de séquences (comme les problèmes de récurrence ou de concurrence des événements), tout en imposant sa grille de lecture et ses normes de scientificité aux travaux mobilisant l'AO. Certains utilisateurs de l'AO prennent ainsi maintes précautions - en s'excusant de produire des analyses « simplement » descriptives, en surévaluant le manque de robustesse de leurs résultats ou en justifiant leurs choix exclusivement à l'aide d'indicateurs statistiques sur le modèle des tests stochastiques - qui ne sont utiles et pertinentes que dans le cadre épistémologique du positivisme tel que défini par Bryant (voir note 21), dans lequel s'inscrit parfaitement l'EHA²³. Ce phénomène se présente comme une nouvelle différenciation fractale, entre l'analyse de séquences « pure » et une forme d'AO adaptée aux standards de la réalité linéaire générale²⁴. On peut alors se demander si la sécularisation de l'AO, *i.e.* son autonomisation par rapport aux travaux d'Abbott, ne s'opère pas au prix d'un relatif dévoiement de l'approche théorique et méthodologique à laquelle elle était initialement associée, ou bien s'il s'agit uniquement de l'implacable normalisation de l'usage de la méthode²⁵.

²² Rosental parle également de « laïcisation » à propos de l'analyse factorielle (Rosental, 1997).

²³ Brendan Halpin et Tak Wing Chan, tous deux auteurs de plusieurs travaux utilisant l'AO, écrivent ainsi en 1998 : « Nous sommes prédisposés (*sic*) à être méfiants vis-à-vis des boîtes noires, de la classification en général, et des affirmations qui présentent l'AO et les procédures apparentées comme un mouvement des analyses centrées sur les variables vers celles centrées sur les récits » (« *We are predisposed to be wary about black boxes, of cluster analysis in general, and of claims that OMA and cognate procedures represent a movement from variable-centred to narrative-centred analyses* ») ; ils concluent : « L'AO aurait plus de valeur si l'on pouvait l'intégrer directement dans une modélisation statistique » (« *it would enhance the value OMA if it could be linked directly into statistical modelling* »). De même, Chan en 1999 : « il me semble que cela se rapproche terriblement de l'utilisation de l'AO comme une machine à faire les saucisses » (« *It seems to me that this comes perilously close to using OMA as a sausage machine* »). On notera que ces citations, extraites de leur contexte, ne sont en aucun cas représentatives du positionnement général de leurs auteurs. Elles ont été choisies avant tout pour leur caractère emblématique du type de propos que l'on trouve disséminés dans de nombreux travaux. Un cas limite consiste en la réorientation de chercheurs d'une branche de la fractale à l'autre. Ainsi McVicar et Anyadike-Danes réalisent dans les années 2000 trois analyses de séquences d'insertion professionnelle. La stratégie méthodologique est très rigoureuse et les auteurs prennent soin de démontrer la robustesse des résultats. Pourtant, en 2010, ils publient un article en forme de palinodie, intitulé « *Does Optimal Matching Really Give Us Anything Extra for the Analysis of Careers Data?* » (« l'AO apporte-t-il quelque chose en plus pour l'analyse des carrières ? »), et concluent : « Si l'on accepte le principe selon lequel les carrières sont le résultat d'un processus stochastique (*sic*), alors l'AO ne comble aucune lacune méthodologique majeure » (« *if we accept careers are generated stochastically through time then there are no major methodological 'gaps' that OM fills in this context* »).

²⁴ On observe un phénomène analogue avec l'analyse des réseaux sociaux, entre les tenants de l'analyse structurale « pure » et ceux d'une approche plus « souple » des réseaux. Par ailleurs, une autre hypothèse - compatible - est plausible, celle d'une cristallisation d'un mode d'appropriation unique de la méthode (Blanchard, 2010, p. 11), à mesure que les pratiques se normalisent et que les usages « innovateurs » et « technicistes » se raréfient.

²⁵ Un indice de cette « routinisation du charisme » (Weber, [1921] 1995) d'Abbott parmi les adeptes de l'AO pourrait être la publication, inédite à ce jour, d'articles utilisant l'AO sans mentionner une seule référence aux travaux d'Abbott.

Bibliographie

Abbott, A. 1992. From Causes to Events. Notes on narrative positivism. *Sociological Methods & Research* 20 (4): 428-455.

Abbott, A. 1997. On the concept of turning point. *Comparative Social Research* 16: 85-106.

Abbott, A. 2000. Reply to Levine and Wu. *Sociological Methods & Research* 29 (1): 65-76.

Abbott, A. 2001a. *Chaos of disciplines*. The University of Chicago Press.

Abbott, A. 2001b. *Time matters. On theory and method*. The University of Chicago Press.

Abbott, A. 2004. *Methods of Discovery: Heuristics for the Social Sciences*. WW Norton & Co.

Abbott, A., A. Hrycak 1990. Measuring resemblance in sequence data: an optimal matching analysis of musicians' careers. *American journal of sociology* (96): 144-185.

Abbott, A., A. Tsay 2000. Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological methods & research* 29 (1): 3-33.

Abbott, A., J. Forrest 1986. Optimal Matching Methods for Historical Sequences. *Journal of Interdisciplinary History* 16 (3): 471-494.

Ben-David, J., R. Collins 1966. Social factors in the origins of a new science: the case of psychology. *American Sociological Review* 31 (4): 451-465.

Blanchard, P. 2010. *Analyse séquentielle et carrières militantes*. Rapport de recherche [<http://hal.archives-ouvertes.fr/hal-00476193>]

Bourdieu, P. 1975. Méthode scientifique et hiérarchie sociale des objets. *Actes de la Recherche en Sciences Sociales* 1 : 4-6.

Bourdieu, P. 2001. *Science de la science et réflexivité*. Raisons d'agir.

Bourdieu, P., J.-C. Passeron 1964. *Les héritiers : les étudiants et la culture*. Paris: Ed. de minuit.

Bryant, C. 1989. Le positivisme instrumental dans la sociologie américaine. *Actes de la Recherche en Sciences Sociales* 78 : 64-74.

Chan, T. W. 1995. Optimal matching analysis: a methodological note on studying career mobility. *Work and occupations* 22 (4): 467-490.

Chan, T. W. 1999. Optimal matching analysis. *Social Research Update* (24).

Cibois, P. 1981. Analyse des données et sociologie. *L'Année sociologique* 31: 333-348.

Clogg, C. C. 1992. The Impact of Sociological Methodology on Statistical Methodology. *Statistical Science* 7 (2): 183-196.

Crane, D. 1969. Social structure in a group of social scientists: a test of the "invisible college" hypothesis. *American Sociological Review* 34 (3): 335-352.

Desrosières, A. 2001. Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative. *Genèses* (43): 112-127.

Desrosières, A. 2008. Analyse des données et sciences humaines: comment cartographier le monde social ?. *Journ@l électronique d'histoire des probabilités et de la statistique* 4 (2): 11-19.

- Fabiani, J.-L. 2003. Pour en finir avec la réalité unilinéaire. Le parcours méthodologique de Andrew Abbott. *Annales. Histoire, Sciences Sociales* 58 (3): 549-565.
- Freeman, L. 2004. *The development of social network analysis. A study in the sociology of science*. Vancouver: Empirical Press.
- Gabadinho, A., G. Ritschard, N. Müller, M. Studer 2011. Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software* 40 (4): 37.
- Halpin, B. 2010. Optimal Matching Analysis and Life-Course Data: The Importance of Duration. *Sociological Methods & Research* 38 (3): 365-388.
- Halpin, B., T. W. Chan 1998. Class careers as sequences: an optimal matching analysis of work-life histories. *European Sociological Review* 14 (2): 111-130.
- Heilbron, J. 2003. A regime of disciplines. Toward a historical sociology of disciplinary knowledge. In *The dialogical turn. Roles for sociology in the post disciplinary age*, ed. C. Camic and H. Joas, 23-42. Lanham (MD): Rowman and Littlefield.
- Kadushin C. 1966. The friends and supporters of psychotherapy: on social circles in urban life. *American Sociological Review* 31 (6): 786-802.
- Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. New ed of 3 Revised ed: University of Chicago Press.
- Lamont, M. 1987. How to become a dominant French philosopher: the case of Jacques Derrida. *American Journal of Sociology* 93 (3): 584-622.
- Le Roux, B., H. Rouanet 2011. *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*. 1st ed. Softcover of orig. ed. 2004: Springer.
- Lesnard, L. 2010. Setting Cost in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns. *Sociological Methods & Research* 38 (3): 389-419.
- Lesnard, L., T. de Saint Pol 2009. Décrire des données séquentielles en sciences sociales : panorama des méthodes existantes. Paper presented at the meeting of the Xe Journées de Méthodologie Statistique, Paris.
- Levine, J. H. 2000. But what have you done for us lately? Commentary on Abbott and Tsay. *Sociological methods & research* 29 (1): 34-40.
- Macindoe, H., A. Abbott 2004. Sequence analysis and optimal matching techniques for social science data. In *Handbook of Data Analysis*, ed. M. Hardy and A. Bryman, 387-406. London: Sage.
- Martin, O. 2002. Mathématiques et sciences sociales au XXème siècle. *Revue d'Histoire des Sciences Humaines* 6(1): 3-13.
- McVicar, Duncan and Anyadike-Danes, Michael 2010. Does optimal matching really give us anything extra for the analysis of careers data ? An application to British criminal careers. *Working paper*.
- Merton, R. K. 1968. The Matthew Effect in Science. *Science* 159 (3810): 56-63.
- Mills, C. W. 1959. *The Sociological Imagination*. Oxford University Press.
- Ollion, E. 2011. De la sociologie en Amérique. Eléments pour une sociologie de la sociologie étasunienne contemporaine. *Sociologie* 3 (3): 277-294.
- Passeron, J.-C. 1990. Biographies, flux, itinéraires, trajectoires. *Revue française de sociologie* 31 (1): 3-22.

- Price, D. J. de S. & Beaver, D. 1966. Collaboration in an invisible college. *American Psychologist* 21(11) : 1011-1018.
- Raftery, A. E. 2001. Statistics in sociology, 1950-2000: A selective review. *Sociological methodology* 31: 1-45.
- Robette, N. 2011. *Explorer et décrire les parcours de vie : les typologies de trajectoires*. Paris: CEPED.
- Robette, N. 2012. Harpoon or maggot ? A comparison of various metrics to fish for life course patterns. *Bulletin de Méthodologie Sociologique*. À paraître
- Rohwer, G., U. Pötter 2005. *TDA's user manual*.
- Rosental, P.-A. 1997. Outil ou fétiche: la laïcisation de l'analyse factorielle dans les sciences sociales. *Histoire & Mesure* XII (3/4): 185-196.
- Rouanet, H. 2008. Des statistiques à la Statistique. *site personnel* [<http://www.mi.parisdescartes.fr/~lerb/rouanet/index.html>].
- Savage, M. 2009. Contemporary Sociology and the Challenge of Descriptive Assemblage. *European Journal of Social Theory* 12 (1): 155-174.
- Scherer, S. 2001. Early career patterns: a comparison of Great Britain and West Germany. *European Sociological Review* 17 (2): 119-144.
- Tuma, N. B., M. T. Hannan, L. P. Groeneveld 1979. Dynamic Analysis of Event Histories. *American Journal of Sociology* 84 (4): 820-854.
- Van der Heijden, P. G. M. 1987. *Correspondence analysis of longitudinal categorical data*. Leiden: DSWO Press.
- van Meter, K., M.-A. Schiltz, P. Cibois, L. Mounier 1994. Correspondence analysis: A history and French sociological perspective. In *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*, ed. M. Greenacre and J. Blasius, 128-137. San Diego: Academic Press.
- Weber, M. [1921] 1995. *Économie et société/1*. Paris : Plon-Presses -Pocket.
- Wu, L. L. 2000. Some comments on "Sequence analysis and optimal matching methods in sociology: Review and prospect". *Sociological methods & research* 29 (1): 41-64.