



Analyse harmonique qualitative ou méthodes d'appariement optimal? Une analyse exploratoire de trajectoires professionnelles

Nicolas Robette, Nicolas Thibault

► To cite this version:

Nicolas Robette, Nicolas Thibault. Analyse harmonique qualitative ou méthodes d'appariement optimal? Une analyse exploratoire de trajectoires professionnelles. Population (édition française), 2008, 63 (4), pp.621-646. halshs-01016116

HAL Id: halshs-01016116

<https://shs.hal.science/halshs-01016116>

Submitted on 27 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



NICOLAS ROBETTE* ET NICOLAS THIBAUT**

Analyse harmonique qualitative ou méthodes d'appariement optimal ?

Une analyse exploratoire de trajectoires professionnelles

Les parcours de vie modélisés par les interactions entre les différents domaines d'implication des individus constituent des données complexes dont l'analyse quantitative reste un défi. Pour aborder la complexité biographique, une possibilité est de s'intéresser à certaines transitions spécifiques qui jalonnent les trajectoires. Une autre possibilité consiste à décrire l'ensemble du parcours. Passer de l'examen de transitions à la description typologique de l'ensemble des parcours est un changement de perspective très avantageux, qui conduit du particulier à l'ensemble. C'est ce que Nicolas ROBETTE et Nicolas THIBAUT proposent ici. Confrontant l'analyse harmonique qualitative issue de la tradition française d'analyse des données et l'appariement optimal issu de la génétique, les auteurs fournissent une comparaison bienvenue de ces deux approches alors que ces méthodes se diffusent.

Les enquêtes biographiques, sur le modèle de l'enquête *Triple biographie* réalisée par l'Ined en 1981, collectent de manière rétrospective les trajectoires complètes d'individus, le plus souvent année par année. Ainsi, ces données appellent une description statistique des parcours de vie. L'enquête *Biographies et entourage* de l'Ined a été réalisée en 2001 auprès de 2 830 individus nés de 1930 à 1950, représentatifs de la population francilienne des générations considérées à la date d'enquête (Lelièvre et Vivier, 2001). Elle avait pour objectif principal d'aider à la compréhension de la mobilité résidentielle et familiale des enquêtés en liaison avec celle de leur entourage. Toutefois, le volet professionnel contient des données suffisamment intéressantes pour mériter une exploitation spécifique. Les différentes professions occupées par chaque enquêté au cours de sa vie ont été relevées selon un calendrier rétrospectif de dimension

* Institut national d'études démographiques, Paris.

** Au moment de la rédaction de cet article, Nicolas Thibault travaillait à l'Institut national d'études démographiques.

annuelle. Chaque étape a ensuite été codée selon la nomenclature des PCS de l'Insee. Si les trajectoires professionnelles peuvent en théorie se répartir sur tout le champ des possibles, elles n'en dessinent pas moins de grandes tendances qui sont liées aux mutations de la société (Marchand et Thélot, 1997).

L'objectif de cet article est de déterminer la meilleure façon de traiter ces trajectoires que nous qualifierons de complexes au sens où (i) tous les états qui les caractérisent peuvent se reproduire au cours du temps et (ii) il existe de multiples transitions⁽¹⁾ possibles entre ces états. Une possibilité conforme au paradigme de l'analyse des biographies (Courgeau et Lelièvre, 1996) est d'utiliser des méthodes de statistique exploratoire pour dégager des typologies de parcours, qui appréhendent les trajectoires individuelles dans leur globalité et non plus simplement sous l'angle des événements qui les composent (Billari, 2001). Différentes méthodes typologiques existent (Grelet, 2002) – distance du khi², distance euclidienne (Espinasse, 1993), indicateurs synthétiques, etc. Deux d'entre elles émergent dans la littérature comme particulièrement adaptées à ce type de données, c'est-à-dire capables de décrire de manière systématique la séquence des événements et leur durée. D'une part, l'analyse harmonique qualitative (AHQ), mise au point par des statisticiens français au tournant des années 1980, est une méthode d'analyse de données qui tient compte du temps. D'autre part, les méthodes d'appariement optimal⁽²⁾ (MAO) sont des techniques algorithmiques importées des sciences de la vie par des sociologues américains dans la seconde moitié des années 1980. Nous nous proposons ici de discuter l'intérêt comparé de ces deux techniques à partir de l'exemple des parcours professionnels des hommes retracés dans l'enquête *Biographies et entourage* (Ined, 2001).

I. La nécessité d'une analyse exploratoire des parcours professionnels

1. Le traitement des trajectoires complexes

Classiquement, différents outils statistiques permettent de s'intéresser aux durées de séjour dans un état considéré : on utilise des estimations non-paramétriques pour mesurer les durées de séjour (Kaplan et Meier, 1958 ; Nelson, 1972 ; Aalen, 1978) et on met en œuvre des modèles semi-paramétriques (Cox, 1972) ou paramétriques⁽³⁾ pour mesurer l'impact de caractéristiques individuelles (qui n'en sont pas moins des propriétés sociales) sur la durée de séjour dans un état. Ces modèles sont pertinents dès lors que l'on mesure une durée bien définie par une date de début et une date de fin non problématiques.

(1) Par exemple, dans notre travail, il existe 9 états, et il y a donc 81 transitions possibles.

(2) Traduction de l'expression anglaise, *Optimal Matching Analysis* (OMA) proposé par Laurent Leosnard et Thibaut de Saint-Pol (2004).

(3) Les modèles de durée paramétriques supposent que le séjour dans un état considéré suit une loi déterminée, qui est elle-même fonction du temps.

Cependant, ils n'ont pas pour finalité de décrire des parcours individuels caractérisés par une chronologie complexe de changements d'état c'est-à-dire lorsque l'on étudie une succession d'événements renouvelables et qu'il existe de multiples transitions possibles entre les états au cours de la trajectoire (GRAB, 2006), comme c'est par exemple le cas des parcours professionnels. Se pose alors la question de l'exploration de ce type de données.

2. Le problème des événements renouvelables

Les méthodes longitudinales modernes, qui traitent des parcours individuels, ont déjà montré leur intérêt dans l'étude des parcours sur le marché du travail. À partir du fichier historique de l'Agence nationale pour l'emploi (ANPE), on peut par exemple étudier les parcours de réinsertion des chômeurs dans le marché du travail en fonction de caractéristiques individuelles (Degenne et Lebeaux, 1999) et en fonction des prestations et des mesures dont chaque chômeur a pu bénéficier (Crépon, Gurgand et Dejemeppe, 2005). L'utilisation de modèles de durée s'impose puisqu'on estime le temps qui sépare l'enregistrement dans les fichiers de l'ANPE du retour à l'emploi. Cette définition de la durée du chômage peut permettre de mesurer l'efficacité des politiques d'emploi. Néanmoins, elle néglige le fait que ce retour à l'emploi peut n'être que temporaire et aboutir rapidement à une nouvelle période de chômage. Les modèles de durée ne peuvent pas, en effet, prendre en compte directement le caractère renouvelable des événements⁽⁴⁾.

De la même façon, lorsque l'on analyse les interruptions de carrière professionnelle, on définit généralement la durée étudiée comme la différence entre le début de la $n^{\text{ième}}$ période d'inactivité et la date de reprise d'activité suivante. Mais lorsqu'on s'intéresse à la complexité des parcours caractérisés par des intermittences d'activité, l'analyse d'un événement unique n'est pas suffisante, car les individus peuvent connaître plusieurs interruptions de carrière. Divers aménagements ont pu être proposés pour traiter des épisodes répétés. En premier lieu, on peut distinguer les sorties d'activité par rang. La distinction par rang de naissance est une méthode traditionnelle pour l'étude de la fécondité, mais au contraire des enfants, les sorties d'activité n'ont pas un sens particulier selon leur rang ; c'est plutôt leur longueur, voire leur longueur cumulée qui a une influence sur une carrière professionnelle (Desplanques et Saboulin, 1986 ; Lelièvre, 1987 ; Cambois et Lelièvre, 1988). *In fine*, cette méthode n'est pas pleinement satisfaisante car elle ramène la durée des étapes d'inactivité à la première définition (la différence entre le début de la $n^{\text{ième}}$ période d'inactivité et la reprise d'activité suivante). En second lieu, on peut raisonner sur l'ensemble des périodes en considérant les personnes qui ont connu une ou plusieurs périodes d'inactivité comme un niveau d'agrégation

(4) La solution adoptée par Gurgand, Crépon et Dejemeppe (2005) pour répondre à ce problème est de considérer qu'on n'est pas réellement sorti du chômage en deçà d'un certain seuil de temps en emploi.

dans un modèle multiniveau (Courgeau, 2000). Cette piste de recherche reste encore à explorer et à rendre opérationnelle.

3. Le problème de la multiplicité des états et des transitions

Le problème de définition de la durée se pose avec encore plus d'acuité lorsque l'état considéré ne se mesure pas de manière binaire mais peut se caractériser par une multiplicité de situations : appartenance à différentes catégories socioprofessionnelles, travail à temps plein, à temps partiel, chômage, inactivité.

Les démographes cherchent par exemple à mesurer la durée de l'activité et à la mettre en relation avec d'autres événements, comme la naissance des enfants. Ils posent la question classique de l'interaction entre activité, notamment féminine, et fécondité. Les méthodes biographiques semblent plus adaptées pour la traiter que l'analyse par cohorte qui nécessite le recours à l'hypothèse d'indépendance entre événements. Toutefois, elles se heurtent selon M. Kempeneers et É. Lelièvre (1991) à trois problèmes imbriqués, à savoir la définition de l'état dont on estime la durée et des variables exogènes, la définition de la population soumise au risque et la définition de l'intervalle d'étude⁽⁵⁾. Les choix de définition conditionnent la solidité du raisonnement et rendent les conclusions tributaires d'hypothèses de départ, souvent implicites. Les auteures recommandent donc une analyse plus descriptive qui rendrait compte de toutes les activités depuis l'âge de 15 ans. Cela les conduit à préconiser un nouveau type d'étude permettant non seulement de reconstituer les dates de sortie d'activité, mais également de suivre les biographies dans leur intégralité.

II. Construire une typologie de trajectoires professionnelles complexes

1. Les trajectoires étudiées

L'unité statistique sur laquelle porte la classification est la trajectoire individuelle : les classes sont constituées en fonction de la ressemblance entre les

(5) Les auteures traitent de l'interaction entre vie familiale et vie professionnelle des femmes. D'une part, l'analyse suppose souvent l'existence d'une relation univoque : faut-il alors considérer que l'activité a une influence sur la fécondité ou au contraire que la fécondité a une influence sur l'activité ? Le choix de cette problématique de départ n'est aucunement neutre, et les conclusions en dépendent nécessairement. D'autre part, le choix de l'échantillon oblige aussi à poser un certain nombre d'hypothèses fortes. Doit-on n'étudier que les femmes qui ont eu des enfants ? De la même façon, doit-on limiter l'échantillon aux femmes qui sont entrées sur le marché du travail ou considérer que le fait de ne pas y entrer peut relever d'un projet anticipé de fécondité ? Enfin, le choix de l'événement initial est aussi déterminant. Si on retient le projet d'une étude sur la durée d'activité avant interruption, plusieurs événements initiaux sont également légitimes selon la question que l'on se pose ; ils n'en sont pas moins problématiques du fait de la complexité des biographies. Le choix du moment de l'entrée sur le marché du travail comme événement initial néglige les femmes qui ont eu un enfant avant de travailler. Le mariage (ou le début de l'union) pose un problème du même type puisque les enfants peuvent naître avant le début de l'union considérée.

parcours. Un certain nombre de choix sont nécessaires concernant (i) la population étudiée, (ii) l'intervalle d'étude, (iii) la construction des variables de l'analyse.

Nous n'analysons ici que les carrières masculines ($n = 1\,341$). Une classification sur l'intégralité de la population est possible mais elle a deux inconvénients. D'une part, la liste des états est différente pour les hommes et les femmes : le service militaire n'a d'influence que sur les itinéraires masculins. D'autre part, lorsque l'on traite ensemble les deux sexes, les différences de carrières entre hommes et femmes peuvent parfois être masquées par la simplification impliquée par la classification⁽⁶⁾.

Les individus sortent d'observation à la date de l'enquête : les données sont donc censurées à droite⁽⁷⁾. Les modèles de durée permettent de contrôler l'effet des troncatures mais pas les statistiques descriptives. Techniquement, rien n'empêche d'appliquer aussi bien l'analyse harmonique qualitative (Barbary et Pinzon Sarmiento, 1998) que les méthodes d'appariement optimal (Macindoe et Abbott, 2004) à des trajectoires de longueurs différentes. Toutefois, avec une telle procédure, on sort du cadre de la description d'une même population au cours du temps. La description de trajectoires individuelles suppose que tous les individus soient observés sur une période identique, délimitée par les mêmes bornes.

Notre étude porte ainsi sur la mobilité professionnelle entre l'âge de 14 ans, qui marque la fin de la scolarité obligatoire pour les générations étudiées⁽⁸⁾, et celui de 50 ans, qui est l'âge des enquêtés les plus jeunes⁽⁹⁾. On aurait pu continuer l'analyse au-delà de 50 ans en travaillant sur une sous-population de l'enquête. Toutefois, cette démarche aurait une validité limitée car le nombre d'enquêtés décroît rapidement avec l'âge : on ne raisonne plus que sur 67,4 % de la population à 55 ans, 42,5 % à 60 ans, 21,3 % à 65 ans et 4,6 % à 70 ans.

Nous décrivons les différents états constituant les trajectoires à partir des groupes socioprofessionnels, c'est-à-dire la nomenclature des PCS en huit postes. Un découpage trop fin des états n'est pas souhaitable car passer de la catégorie

(6) Une description des parcours à partir de l'ensemble de la population a pourtant l'avantage de montrer les proximités entre certaines carrières masculines et féminines. Mais cela ne remet pas en cause notre parti pris de traiter séparément les hommes et les femmes. Si les carrières sont ressemblantes, on dégagera des types proches, comme c'est le cas chez les cadres. Si les itinéraires sont dissemblables, on évite de lisser les disparités liées au sexe.

(7) Par construction, les enquêtes étant rétrospectives, la date d'enquête est la seule modalité de censure à droite : les individus décédés ou ayant migré ne sont pas observés. Le relevé a été suffisamment exhaustif pour qu'il n'y ait pas de censure à gauche, c'est-à-dire présence d'individus dont la biographie ne serait connue qu'à partir d'une date donnée.

(8) Certains enquêtés ont déclaré des étapes professionnelles dès l'âge de huit ans, par exemple parce qu'ils étaient bergers dans une ferme et ne se rendaient qu'épisodiquement à l'école. Nous négligerons ces étapes dans l'analyse car la consigne donnée aux enquêtés faisait commencer la datation à 14 ans. Ce choix est cohérent avec les principes de la nomenclature des PCS qui ne s'applique qu'aux personnes âgées de plus de 15 ans.

(9) On raisonne en âge atteint (en différence d'années) et non en âge exact. Nous connaissons l'âge exact au moment de l'enquête mais les événements biographiques sont connus en âge atteint.

employé administratif à celle d'employé de commerce n'a pas le même sens que de devenir cadre. Par construction, l'échantillon ne contient pas de retraité puisque la description s'arrête à 50 ans. On aurait donc dû raisonner sur sept groupes socioprofessionnels, à savoir les six groupes d'actifs et les « autres personnes sans activité professionnelle ». Toutefois, il nous est apparu intéressant de scinder cette dernière catégorie entre les étudiants et ceux qui sont sans activité pour des raisons familiales ou de santé⁽¹⁰⁾. On a ajouté un dernier groupe issu de la nomenclature de 1954 (mais aujourd'hui supprimé) : celui des militaires du contingent, qui s'avère nécessaire pour restituer les parcours masculins pour ces générations qui ont notamment connu la guerre d'Algérie.

2. Constitution d'une typologie par analyse harmonique qualitative

L'analyse harmonique est une branche des mathématiques qui a connu de nombreuses applications en sciences physiques ou en biologie. Son utilisation en sciences sociales est plus récente et date des années 1970 (Deville, 1974 ; 1977). Il s'agissait alors d'introduire la durée dans l'explication des phénomènes sociaux grâce à des données sur les parcours individuels. Comme le souligne Deville (1977, p. 18) : « Devant des données d'une telle richesse, le statisticien éprouve une certaine perplexité. Des tableaux de plus en plus complexes deviennent ininterprétables sans le secours de méthodes d'analyse "automatique". Il cherche alors à définir une méthode d'analyse qui lui permette de tirer l'essentiel des données dont il dispose. Le mot "essentiel" prend alors un sens précis, quantifiable, lié à la méthode qu'il met en œuvre. » Cette technique a ensuite été adaptée pour en faire une technique de statistique exploratoire des trajectoires complexes (Deville et Saporta, 1980 ; Deville, 1982) appelée analyse harmonique qualitative.

Celle-ci consiste à déterminer une période d'observation, à la diviser en un nombre fini d'intervalles puis à mesurer pour chaque individu la proportion du temps passé dans chacun des états dans chaque intervalle. Une analyse factorielle des correspondances sur la matrice⁽¹¹⁾ ainsi constituée permet de résumer l'information en sélectionnant les facteurs portant le plus d'inertie (Deville, 1982). On élimine de cette façon le « bruit » statistique sans éliminer d'individu. Les progrès des méthodes de classification ont permis, entre autres possibilités, d'utiliser une classification ascendante hiérarchique pour déterminer des typologies de parcours (Barbary, 1996 ; Degenne, Lebeaux et Mounier, 1996 ; Barbary et Pinzon Sarmiento, 1998). La typologie ainsi construite tient compte, d'une part, de la succession des états, du moment auquel les états

(10) La catégorie inclut aussi les chômeurs n'ayant jamais travaillé (qui sont actifs au sens de l'Insee), mais ce cas peut être négligé puisqu'il ne concerne que trois étapes sur les 19 930 étapes de l'enquête. Ce faible nombre est la conséquence du caractère annuel des données.

(11) Avec en lignes les individus et en colonnes le temps passé dans chacun des états pendant chaque intervalle.

interviennent mais aussi des durées passées dans les différents états : elle permet, d'autre part, de conserver tous les individus dans l'analyse.

L'analyse harmonique qualitative a été peu utilisée jusqu'à présent, du fait du manque de données adaptées. Ainsi, J.-C. Deville (1982) constitue un échantillon *ad hoc* à partir des questionnaires éliminés lors des enquêtes Famille de l'Insee de 1962 et 1975⁽¹²⁾. L'échantillon n'est par conséquent aucunement représentatif et a pour seule ambition de fournir un terrain d'application à une méthode qui n'en avait pas. Sur plus de vingt ans, la méthode n'est utilisée qu'une fois sur des données françaises⁽¹³⁾ (Degenne, Lebeaux et Mounier, 1996). Ces travaux portent sur une partie du parcours professionnel pour étudier l'insertion dans le marché du travail. Finalement, il faut attendre le renouveau des collectes biographiques pour voir des applications se concrétiser sur des données latino-américaines (Dureau *et al.*, 1994 ; Barbary, 1997 ; Barbary et Pinzon Sarmiento, 1998). L'enquête *Biographies et entourage* permet de suivre les parcours migratoires par cette méthode (Bonvalet, Bringé et Robette, 2009) mais aussi l'intégralité des carrières professionnelles⁽¹⁴⁾.

La constitution de la matrice harmonique

Une fois la période d'étude fixée, on doit la découper en intervalles pour l'analyse. Les données étant de dimension annuelle, le découpage en périodes d'un an peut sembler le plus naturel et le plus apte à restituer l'information biographique. Toutefois, une telle procédure n'est pas optimale car elle engendre un tableau dont la grande majorité des cases sont nulles, ce qui hypothèque la qualité de l'analyse⁽¹⁵⁾. À l'opposé, un nombre trop réduit d'intervalles entraînerait la perte d'une partie de la richesse de l'information disponible. Il y a donc un arbitrage à effectuer pour fixer le nombre des intervalles.

(12) L'échantillon est constitué d'un certain nombre de femmes mariées trois fois ou plus dont l'histoire conjugale avait été considérée comme trop complexe pour être traitée.

(13) Signalons qu'A. Degenne, M.-O. Lebeaux et L. Mounier (1996) introduisent de surcroît dans la matrice des variables représentant les transitions d'un état à un autre : les neuf groupes que nous distinguons engendreraient quatre-vingt-un types de transitions. En plus du temps passé dans chaque état à chaque période, on reporte pour chaque individu le nombre de transitions professionnelles de chaque type. Ici, comme il s'agit de comparer nos deux méthodes, nous nous en tenons à l'analyse harmonique qualitative au sens strict du terme.

(14) En 1989, l'enquête *Carrière et mobilité* de l'Insee avait procédé à un relevé biographique incomplet : on demandait la profession occupée à cinq dates différentes entre 1960 et 1989. Elle permettait donc de reconstituer partiellement les carrières mais pas de suivre les individus de manière continue tout au long de leur parcours professionnel. Elle a toutefois donné lieu à la réalisation d'une typologie des parcours par la méthode des nuées dynamiques, à partir des informations sur la mobilité entre deux dates successives (Goux, 1991). Mais dans la mesure où l'on ne dispose que des professions occupées (ou de l'inactivité) à des dates fixes, le problème du traitement de la durée reste posé.

(15) On observe 1 341 individus pouvant connaître 9 états sur p périodes. La matrice étudiée aura 1 341 lignes. Si on choisit une périodisation annuelle entre 14 et 50 ans, la matrice aura $9 \times 37 = 333$ colonnes. Par construction, l'individu ne connaît qu'un état par an : chaque ligne comporte donc 37 chiffres « 1 » et 296 « 0 ». Sur les $1\,341 \times 333 = 446\,553$ cellules du tableau, seulement un neuvième seront non nulles.

Un autre arbitrage concerne l'amplitude des intervalles. Rien n'oblige à ce que les amplitudes soient égales ; bien au contraire, certains moments de la vie, le plus souvent la jeunesse, sont caractérisés par un nombre important de changements professionnels, d'autres par une mobilité plus faible. Le choix d'intervalles trop petits est d'autant plus gênant que les changements sont nombreux, car deux trajectoires identiques mais légèrement décalées dans le temps pourraient alors être considérées comme éloignées. La taille de l'intervalle doit finalement être assez grande par rapport au rythme global des changements d'état (Florette, 1988). Nos différents tests nous ont amenés à retenir un découpage en dix intervalles correspondant aux déciles de la distribution des changements d'activité en fonction de l'âge⁽¹⁶⁾.

Pour chaque individu, on calcule la proportion de la durée de chaque intervalle passée dans chacun des états possibles⁽¹⁷⁾. On effectue ensuite une analyse factorielle des correspondances de la matrice obtenue, puis une classification ascendante hiérarchique à partir des 25 premiers facteurs qui portent 70 % de l'inertie. Cela permet de réduire l'hétérogénéité des données tout en conservant une part essentielle de l'information⁽¹⁸⁾.

3. Constitution d'une typologie par appariement optimal

La méthode d'appariement optimal s'appuie sur un ensemble d'algorithmes dynamiques utilisés principalement en biologie moléculaire pour analyser les similarités entre chaînes d'ADN. Elle a ensuite été introduite dans les sciences sociales par Andrew Abbott dans les années 1980 (Abbott et Forrest, 1986 ; Abbott et Hrycak, 1990). Son principe est fondé sur la mesure de la similarité ou de la dissimilarité entre des paires de séquences. L'idée de base consiste à mesurer la dissimilarité entre deux séquences en évaluant le « coût » représenté par la transformation de l'une des séquences en l'autre. La transformation est effectuée au moyen d'opérations : l'insertion (un élément est inséré dans la séquence), la suppression (un élément est supprimé de la séquence) et la substitution (un élément est substitué à un autre). Dans la pratique, l'appariement optimal ne repose que sur deux opérations élémentaires. D'une part, l'insertion et la suppression se voient attribuer un même coût, de façon à ce que passer d'une séquence 1 à une séquence 2 par insertion d'un élément soit équivalent à passer d'une séquence 2 à une séquence 1 par suppression d'un élément⁽¹⁹⁾ ; cette opération est donc appelée *indel* par contraction des termes anglais *inser-*

(16) On ne peut évidemment choisir que les valeurs annuelles les plus proches des déciles théoriques. Les déciles sont ainsi fixés à 18, 20, 22, 24, 26, 29, 33, 38 et 43 ans.

(17) On a dix intervalles et neuf états : on crée donc quatre-vingt-dix variables d'état. La matrice finale comporte 1 341 lignes et 90 colonnes.

(18) Les tests effectués montrent que les classes de la typologie obtenue sont relativement stables lorsque l'on conserve entre 50 % et 95 % de l'inertie. Peu d'individus changent de classe quand on modifie la part de l'information retenue.

(19) Par exemple, la différence entre les séquences ABA et AA est annulable indifféremment par la suppression de B dans ABA ou l'insertion de B dans AA.

tion et deletion. D'autre part, la substitution est la combinaison d'une insertion et d'une délétion⁽²⁰⁾ mais rien n'oblige à fixer le coût d'une substitution au double du coût *indel* : le choix du coût de substitution est un arbitrage qui privilégie soit le rapprochement de suites d'éléments identiques mais décalées dans le temps, soit celui de suites concomitantes d'éléments mais comportant une ou plusieurs différences.

Une série d'opérations a un coût équivalent à la somme des coûts des opérations élémentaires. La distance entre deux séquences est alors définie comme le coût minimal de la transformation d'une séquence en l'autre. Des algorithmes dynamiques spécifiques garantissent l'obtention du coût minimal (Sankoff et Kruskal, 1983). L'appariement de l'ensemble des paires de séquences permet la création d'une matrice de distances, que l'on peut ensuite utiliser pour regrouper les séquences les plus similaires, au moyen de méthodes de classification par exemple, et obtenir une typologie. Le choix des coûts des opérations élémentaires constitue une étape essentielle des techniques d'appariement optimal. C'est la possibilité de détermination des coûts qui confère à la méthode sa souplesse et sa capacité à s'adapter à l'objet étudié (Lesnard et Saint-Pol, 2004).

La constitution de la matrice de coûts

Nous travaillons sur les séquences constituées par les professions occupées par les enquêtés sur des périodes annuelles. Il est possible de différencier les coûts de substitution selon les éléments substitués l'un à l'autre. Si certains chercheurs préfèrent adopter des coûts de substitution fixes du fait du manque d'éléments théoriques sur le sujet (Dijkstra et Taris, 1995), de nombreux travaux adoptent des coûts de substitution différenciés selon des hypothèses propres à l'objet étudié : plus les éléments sont similaires, plus le coût de substitution choisi pour l'analyse est faible. Ainsi certains travaux sur les carrières professionnelles fixent les coûts de substitution en fonction des positions relatives des professions au sein d'une hiérarchie (Stovel *et al.*, 1996 ; Halpin et Chan, 1998 ; Blair-Loy, 1999 ; Scherer, 2001 ; Solis et Billari, 2002). Une solution alternative consiste à dériver les coûts de substitution des probabilités de transition entre les éléments : le coût de substitution entre deux éléments est d'autant plus élevé que la probabilité de transition entre ces éléments est faible (Rohwer et Pötter, 2005). Dans la mesure où il n'existe pas de hiérarchie *a priori* entre les catégories socioprofessionnelles, nous avons adopté la seconde solution (annexe 1).

On s'intéresse ensuite à la relation entre coût de substitution et coût *indel*. Selon la littérature, plusieurs possibilités s'offrent à nous. Tout d'abord, une substitution étant équivalente à la combinaison d'une insertion et d'une suppression, le coût *indel* peut être fixé à la moitié du coût de substitution. Ensuite,

(20) Par exemple, la transformation de la séquence ABA en AAA correspond à une suppression de B puis à une insertion de A.

nous pouvons fixer le coût *indel* à une valeur légèrement supérieure à la moitié du coût de substitution maximal, ce qui limite l'utilisation des insertions-suppressions. Cette approche se justifie dans le cas où on privilégie la position des éléments au sein de la séquence (autrement dit, le moment auquel interviennent les événements) plutôt que l'ordre respectif de ces mêmes éléments. Enfin, si l'on privilégie la succession des éléments, un coût *indel* égal à 1/10 du coût de substitution maximal paraît plus approprié (Macindoe et Abbott, 2004). Dans un itinéraire de mobilité intra-générationnelle, l'ordre des différents états est fondamental. Nous avons donc privilégié la deuxième option (annexe 1).

Une illustration du codage des trajectoires avec chacune des deux méthodes est présentée dans l'encadré 1.

ENCADRÉ 1

Représentation des trajectoires avec la MAO et l'AHQ

Pour l'exemple, considérons 7 années d'une carrière (de 14 à 20 ans) : à 14 et 15 ans, l'enquête est étudiant ; entre 16 et 18 ans, il est ouvrier ; enfin, à 19 et 20 ans, il est cadre. Pour simplifier l'exposé, nous nous limiterons à ces trois états (respectivement notés E, O et C). Ce parcours est représenté sous deux formes différentes par les deux méthodes considérées.

Avec l'AHQ, si on découpe la période en deux intervalles, un de quatre ans (14-17 ans) et un de trois ans (18-20 ans), on obtient la matrice suivante :

14-17 ans			18-20 ans		
E	O	C	E	O	C
0,5	0,5	0	0	0,33	0,67

Avec la MAO, on obtient la séquence : EE000CC dans laquelle chaque lettre représente l'état occupé sur une période annuelle. Le passage à la séquence EEE00CC pourrait être réalisé par l'insertion d'un O et la suppression d'un E, ou par la substitution d'un O à un E. On retient alors la suite d'opérations dont la somme des coûts est minimale.

III. Comparaison des deux techniques sur les itinéraires professionnels

1. Des typologies résultantes proches

En travaillant sur l'intégralité des trajectoires professionnelles masculines, nous avons retenu deux partitions⁽²¹⁾ en six et dix classes de parcours qui semblent un bon compromis entre l'exigence de synthèse de l'information et

(21) Les classes ont été agrégées selon le critère de Ward ; nous n'avons pas réalisé de consolidation de la partition, par exemple par la technique des centres mobiles ou celle des *k-means*, de façon à obtenir différents niveaux de partition d'une même classification et à en faciliter la comparaison : la partition en 6 classes est un regroupement de celle en 10 classes.

celle de présentation de l'hétérogénéité des trajectoires individuelles. Les typologies en 6 classes issues de l'AHQ et de la MAO sont relativement proches, 5 classes ayant des profils très comparables avec des effectifs légèrement différents (tableau 1). L'intitulé des classes présente la caractéristique principale de chacune d'entre elles, afin de rendre plus intelligible leur description⁽²²⁾.

**Tableau 1. Typologies en 6 classes des carrières
des hommes franciliens des générations 1930 à 1950**

n°	Classe (caractéristique principale)	AHQ (%)	MAO (%)
1	Professions intermédiaires	29	27
2	Ouvriers	26	26
3	Cadres	26	26
4	Employés	13	9
5	Vers artisans, commerçants et chefs d'entreprise	5	5
6	De professions intermédiaires à cadres	–	6
7	D'agriculteurs à ouvriers	1	–
	Total	100	100

Champ : 1 341 carrières professionnelles d'hommes des générations 1930-1950 résidant en Île-de-France à la date de l'enquête.
Source : enquête *Biographies et entourage* (Ined, 2001).

**Classe 1. Les professions intermédiaires
(29 % avec l'AHQ, 27 % avec la MAO)**

Les hommes de cette classe sont caractérisés par un passage long dans les professions intermédiaires. On peut toutefois distinguer différents parcours. Une première sous-classe agrège des hommes qui sont passés par de nombreux groupes socioprofessionnels ; elle est quantitativement peu importante. Un second type de parcours est constitué d'hommes qui ont commencé leur carrière en tant qu'employés puis sont entrés relativement tôt dans les professions intermédiaires voire ensuite dans le groupe des cadres. Le même type de carrières s'observe pour des hommes qui ont commencé en tant qu'ouvriers. Ces deux sous-classes s'opposent aux parcours de ceux qui sont passés dans les professions intermédiaires au moment de leur entrée dans le marché du travail, parmi lesquels on peut distinguer ceux qui sont devenus cadres avant 50 ans et les autres. La durée qui sépare la fin des études de l'entrée dans les professions intermédiaires est relativement courte. Il s'agit pour beaucoup d'hommes qui sont entrés dans les catégories moyennes par leurs diplômes et non par mobilité professionnelle.

(22) Une typologie constitue une synthèse de la diversité des trajectoires et chaque classe renferme elle-même une certaine hétérogénéité. Cette hétérogénéité intra-classe peut être illustrée à l'aide de graphiques comme les chronogrammes, présentant la distribution des individus selon les différents états à chaque âge (l'exemple de la typologie en 6 classes issues de la MAO est donné en annexe 2). Un autre type de graphique, couramment appelé « tapis », permet de représenter les trajectoires individuelles, mais il est difficilement lisible sans couleurs.

Classe 2. Les ouvriers (26 % avec l'AHQ et la MAO)

Les hommes qui ont été ouvriers au cours de leur carrière se retrouvent dans plusieurs classes : ils peuvent avoir été ouvriers plus ou moins longtemps et à différents moments de leur parcours. La classe 2 regroupe les hommes qui ont été ouvriers sur une très grande partie de la période d'observation même s'ils ont pu connaître d'autres positions au cours de leur carrière. Certains ouvriers accèdent directement aux catégories intermédiaires en devenant contremaîtres ou agents de maîtrise, ce qui peut correspondre à une promotion. Finalement, le phénomène massif est la faiblesse de la mobilité ascendante.

Classe 3. Les cadres (26 % avec l'AHQ et la MAO)

Cette classe réunit les enquêtés qui sont devenus précocement cadres, en général dès leur premier emploi. Des différences apparaissent en fonction de leur âge à l'entrée dans ce groupe.

Classe 4. Les employés (13 % avec l'AHQ et 9 % avec la MAO)

Le type 4 rassemble des parcours relativement hétérogènes, ce qui ne doit pas être compris comme une faiblesse de la méthode que nous proposons. Bien au contraire, le type 4 ne regroupe que des hommes qui ont en commun leur passage relativement long dans le groupe des employés. Les sous-classes qui apparaissent à d'autres niveaux de partition de la classification hiérarchique montrent les formes différentes que peut prendre ce passage. Apparaît ainsi notamment un groupe composé d'hommes qui deviennent employés après avoir été ouvriers dans leur jeunesse ; un autre itinéraire est celui des hommes qui étaient employés puis ont accédé aux professions intermédiaires et sont même pour certains devenus cadres ; un autre groupe réunit les hommes qui sont employés sur une très grande majorité de la période.

Classe 5. Vers artisans, commerçants et chefs d'entreprise (5 % avec l'AHQ et la MAO)

Cette classe agrège principalement des individus qui sont devenus artisans, commerçants et chefs d'entreprise. Le plus souvent, ils ont été auparavant et pendant une durée variable ouvrier, employé, profession intermédiaire ou cadre. On est donc ici en présence d'une forme de mobilité professionnelle.

Classe 6. De professions intermédiaires à cadres (6 % avec la MAO)

Les individus appartenant à la classe 6 connaissent une mobilité intragénérationnelle ascendante, des professions intermédiaires vers le statut de cadre. Cette classe relativement homogène n'apparaît qu'avec la MAO.

Classe 7. Les agriculteurs devenus ouvriers (1 % avec l'AHQ)

La dernière classe, spécifique à l'AHQ, est composée d'individus qui ont été agriculteurs dans leur jeunesse. En début de trajectoire, la plupart d'entre

eux sont aides familiaux chez leur parents, puis cette proportion décroît très rapidement entre 20 et 30 ans, principalement au profit du statut d'ouvrier.

Les typologies en 6 classes issues de l'AHQ et de la MAO semblent construites à partir de la situation occupée le plus longtemps et, de ce fait, elles ne font émerger les phénomènes de mobilité que de manière marginale (classes 5, 6 et 7). Toutefois, des typologies en 10 classes incitent à nuancer ce constat (tableau 2).

Tableau 2. Typologies en 10 classes des carrières des hommes franciliens des générations 1930 à 1950

n°	Classe (caractéristique principale)	AHQ (%)	MAO (%)
1	Cadres	25	26
2	Professions intermédiaires	22	16
3	Ouvriers	19	19
4	Employés	12	6
5	D'ouvriers vers professions intermédiaires	10	6
6	Vers artisans, commerçants et chefs d'entreprise	5	5
7	D'agriculteurs à ouvriers, avant 25 ans	2	–
8	D'agriculteurs à ouvriers, après 25 ans	1	–
9	Inactivité	1	–
10	Artisans, commerçants et chefs d'entreprise	2	–
11	De professions intermédiaires à cadres	–	6
12	D'ouvriers à professions intermédiaires ou artisans, commerçants et chefs d'entreprise, après 35 ans	–	7
13	D'employés à professions intermédiaires	–	5
14	D'ouvriers à employés	–	2
	Total	100	100

Champ : 1 341 carrières professionnelles d'hommes des générations 1930-1950 résidant en Île-de-France à la date de l'enquête.
Source : enquête *Biographies et entourage* (Ined, 2001).

Les typologies en 10 classes sont relativement proches. En particulier, si leurs effectifs diffèrent parfois, les 6 premières classes (qui représentent au total 93 % des enquêtés avec l'AHQ et 78 % avec la MAO) ont des profils très comparables. Les classes 1 à 4 regroupent des parcours stables de cadres, de professions intermédiaires, d'ouvriers et d'employés. Les classes 5 et 6 correspondent en revanche à des parcours de mobilité : passage d'ouvrier à profession intermédiaire, d'une part, et mobilité vers le statut d'artisan, commerçant et chef d'entreprise, d'autre part.

Les autres classes issues de l'AHQ ont des effectifs faibles et font apparaître des phénomènes résiduels. Les classes 7 et 8 concernent ainsi des individus initialement agriculteurs qui deviennent ensuite ouvriers, la distinction étant fonction de l'âge auquel intervient la transition. La classe 9 présente un intérêt

particulier dans la mesure où elle est la seule à accorder une place prépondérante à l'inactivité : elle est composée d'hommes ayant connu un ou plusieurs épisodes d'inactivité, d'une durée totale significativement plus longue que chez les hommes des autres classes.

Enfin, les classes qui ne se révèlent qu'avec la MAO ont des effectifs plus importants et se rapportent uniquement à des parcours de mobilité : de professions intermédiaires vers cadres ; d'ouvriers vers professions intermédiaires ou artisans, commerçants et chefs d'entreprise (la transition intervenant après 35 ans) ; d'employés vers professions intermédiaires ; d'ouvriers vers employés. La MAO semble donc plus sensible aux transitions, en capturant les processus de mobilité intra-générationnelle dans des classes relativement homogènes.

L'apparition dans les typologies en 10 classes de carrières de mobilité invisibles dans les typologies en 6 classes montre l'importance d'explorer une classification de trajectoires à différents niveaux de partition, pour voir par exemple si la population regroupée au sein d'un type de trajectoire particulier pourrait avec profit se subdiviser en plusieurs types distincts. Une typologie en un nombre de classes élevé permet de mettre au jour un large éventail de parcours mais peut se révéler difficile à caractériser. Outre l'exploration de données complexes, une approche typologique vise en effet aussi à synthétiser la variété des comportements individuels. Le choix d'un nombre de classes est donc le résultat d'un arbitrage entre synthèse et exhaustivité, selon les hypothèses et les questions de recherche. Ces résultats constituent en effet une première analyse des trajectoires professionnelles, qu'on pourra approfondir en croisant différentes caractéristiques individuelles (génération, lieu de naissance, profession du père, etc.) avec le type de carrière de l'enquêté, en utilisant des statistiques descriptives. Des modèles de régression, dans lesquels la classe de trajectoires est incluse comme variable dépendante ou explicative, peuvent aussi être utilisés, avec toutefois des précautions du fait que les relations de causalité s'inscrivent dans la diachronie.

2. La mesure des différences

La proximité entre les typologies construites au moyen des deux techniques d'analyse harmonique qualitative et d'appariement optimal peut être illustrée à l'aide d'une matrice de correspondance, qui croise les répartitions de la population selon deux partitions distinctes. On peut ainsi voir si les individus d'une classe issue de la première partition sont plutôt concentrés dans une seule classe de la seconde partition ou s'ils sont au contraire ventilés entre plusieurs classes. Ici, nous présentons en lignes la classification en 6 classes obtenue par appariement optimal et en colonnes celle qui résulte de l'analyse harmonique qualitative (tableau 3). La matrice de correspondance des typologies en 10 classes est reproduite en annexe 3.

Tableau 3. Matrice de correspondance entre les typologies en 6 classes obtenues par appariement optimal et par analyse harmonique qualitative

		AHQ						
		1	2	3	4	5	6	Total
MAO	1	291	16	4	45	5	1	362
	2	7	320	0	5	13	7	352
	3	17	5	311	18	0	0	351
	4	2	11	1	98	0	5	117
	5	13	0	9	2	49	0	73
	6	59	1	23	2	1	0	86
	Total	389	353	348	170	68	13	1 341
<i>Champ</i> : 1 341 carrières professionnelles d'hommes des générations 1930-1950 résidant en Île-de-France à la date de l'enquête. <i>Source</i> : enquête <i>Biographies et entourage</i> (Ined, 2001).								

La lecture intuitive de la diagonale du tableau montre une forte proximité entre les classifications obtenues par les deux méthodes, à l'exception de la classe 6 dont l'effectif est faible⁽²³⁾. Il est possible de calculer un taux de correspondance, indice synthétisant le degré de similitude entre les deux partitions. Formellement, on fait la somme de l'effectif du mode de chacune des lignes et de chacune des colonnes qu'on rapporte au double de l'effectif total. On calcule ainsi la moyenne de la correspondance de la classification obtenue par MAO à celle obtenue par AHQ et de la classification obtenue par AHQ à celle obtenue par MAO :

$$\sigma = \frac{\sum_i \max_j (n_{ij}) + \sum_j \max_i (n_{ij})}{2N} \tag{1}$$

où n_{ij} représente l'effectif appartenant simultanément à la $i^{\text{ème}}$ classe de la première partition et à la $j^{\text{ème}}$ classe de la seconde partition et N est la population totale.

La correspondance entre les typologies obtenues selon les deux méthodes est de 82 % dans le cas d'une partition à 6 classes et de 75 % dans le cas d'une partition à 10 classes. Les deux techniques donnent donc des résultats grandement convergents, ce qui est une bonne indication de leur solidité.

Toutefois, quelques nuances apparaissent, ce qui nous amène à formuler des hypothèses issues de nos observations sur l'intérêt comparé de chacune des méthodes. La comparaison des deux typologies en six et en dix classes révèle quelques différences, qui peuvent s'expliquer de deux manières.

D'une part, les classes obtenues par MAO sont plus sensibles aux transitions, notamment dans le début de la trajectoire où les changements d'état professionnel sont les plus nombreux. Les classes distinguent ainsi particulièrement les parcours stables des parcours de mobilité ; elles sont de ce point de vue plus homogènes qu'en utilisant l'AHQ. Cela vient du fait que la MAO

(23) 86 pour la méthode d'appariement optimal et 13 pour l'analyse harmonique qualitative.

retient la séquence année par année, alors que l'analyse harmonique qualitative est fondée sur la mesure de la proportion de chaque intervalle de temps passé dans chaque état et n'intègre pas à l'analyse l'ensemble des transitions d'un état à un autre.

D'autre part, l'analyse harmonique qualitative fait apparaître quelques classes d'effectifs plus réduits regroupant des trajectoires très typées, c'est-à-dire différant fortement des autres par la nature et la séquence des états sur tout le parcours. On observe par exemple une classe regroupant des individus ayant vécu des épisodes courts et diffus d'interruption d'activité, ainsi que des parcours comprenant une ou plusieurs étapes dans le statut d'agriculteur. Au contraire, la partition obtenue par MAO tend à agréger ces trajectoires marginales à des classes plus importantes.

3. Intérêt comparé de chacune des méthodes

La possibilité que laisse l'analyse harmonique qualitative de ne retenir pour la classification que les principaux facteurs issus de l'analyse factorielle permet d'éliminer une partie du « bruit » contenu par les trajectoires, autrement dit de faire ressortir les caractéristiques les plus structurantes de l'information, ce qui en facilite l'interprétation. Mais l'analyste n'est pas à même de déterminer quelle partie de l'information statistique est ainsi négligée ; on ne peut que contrôler *a posteriori* quels sont les facteurs délaissés par la procédure. Par ailleurs, le découpage de la période d'étude en intervalles, d'amplitude non nécessairement égale, permet d'insister sur les moments des trajectoires dans lesquels se concentrent les événements qui intéressent l'analyste. Cela peut notamment se révéler pertinent dans le cas des parcours professionnels, pour lesquels la majorité des événements se concentrent avant 30 ans. Ainsi, l'analyse harmonique qualitative permet d'insister particulièrement sur la période d'apparition d'un événement et sur la durée des états. Les professions occupées le plus longtemps ont notamment une influence plus forte sur la typologie que dans l'appariement optimal, ce qui permet d'insister sur la stabilité socioprofessionnelle qui reste aujourd'hui forte (Goux, 1991).

Au contraire, l'analyse par appariement optimal conserve l'intégralité du détail de la séquence au lieu de la simplifier (Lesnard et Saint-Pol, 2004). C'est alors à l'analyste de fixer lui-même les coûts en fonction de ses hypothèses théoriques, ce qui détermine ensuite le rapprochement entre certaines trajectoires plutôt qu'entre certaines autres. La fixation du coût *indel* permet notamment de faire porter l'intérêt plutôt sur le type de transition ou plutôt sur le moment d'apparition d'un événement. Comparée à l'analyse harmonique qualitative, l'analyse par appariement optimal insiste sur la séquence des événements et le type de transition qui apparaissent dans celle-ci. Par conséquent, elle donne plus de poids aux transitions d'une catégorie socioprofessionnelle à une autre, ce qui permet de mettre en évidence des parcours de mobilité intra-générationnelle.

Conclusion

Les enquêtes biographiques doivent permettre d'étudier statistiquement les parcours de vie. Pourtant, les promoteurs de ces enquêtes avouent eux-mêmes que les résultats sont souvent inférieurs aux espérances. Ainsi, dans un livre consacré aux collectes biographiques quantitatives publié par le Groupe de réflexion sur l'approche biographique, nous pouvons lire que « [...] nous sommes en deçà de l'utilisation potentielle de telles données. On peut dire que l'effort énorme qui a été développé pour produire ces données et acquérir les compétences nécessaires pour les traiter de façon adéquate pouvait et devait produire davantage, avoir une application et une diffusion très large » (GRAB, 1999, p. 46, citation de M. Bottai). En effet, les questionnaires sont assez longs, ce qui rend la collecte coûteuse. Les échantillons sont nécessairement limités à quelques milliers d'individus. L'information est par conséquent difficile à traiter car elle retrace une multiplicité d'événements relatifs à un petit nombre d'individus.

Les méthodes typologiques d'exploration de trajectoires pallient cette difficulté en fournissant un outil de traitement permettant de restituer quantitativement la logique des parcours individuels. Elles sont ainsi adaptées à ces sources de données prometteuses pour les démographes et plus généralement pour les sciences sociales. Encore faut-il se repérer parmi les différentes méthodes possibles, dont certaines sont encore peu connues et dont les avantages comparés sont rarement testés. Il ressort de notre travail que l'analyse harmonique qualitative semble plus adaptée si on cherche à insister sur la durée dans certaines étapes et donc à mettre l'accent sur les états dans lesquels l'individu reste le plus longtemps. Au contraire, l'analyse par appariement optimal a un avantage si l'on cherche à rapprocher les trajectoires sur la base du type de transition et insister ainsi sur les parcours de mobilité. Mais les données longitudinales individuelles sont fortement structurées et les principaux types de parcours apparaissent quelle que soit la méthode utilisée.

Remerciements. Nous tenons à remercier Valérie Golaz, Maryse Marpsat et Thibaut de Saint-Pol pour leur lecture de la première version de cet article. La présente version doit beaucoup à leurs commentaires. Nous remercions également Vincent Cardon pour sa relecture de la présente version.



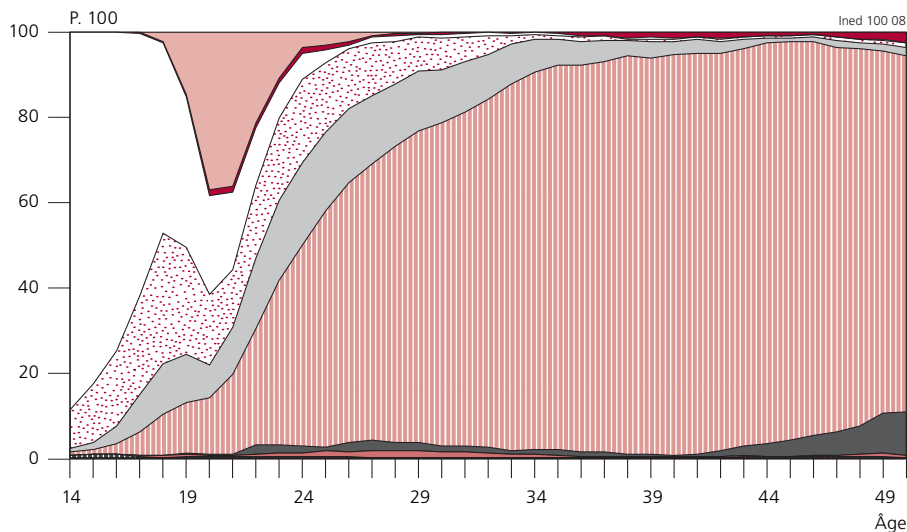
ANNEXES

Annexe 1. Matrice des coûts de substitution
et coût *indel** de l'appariement optimal

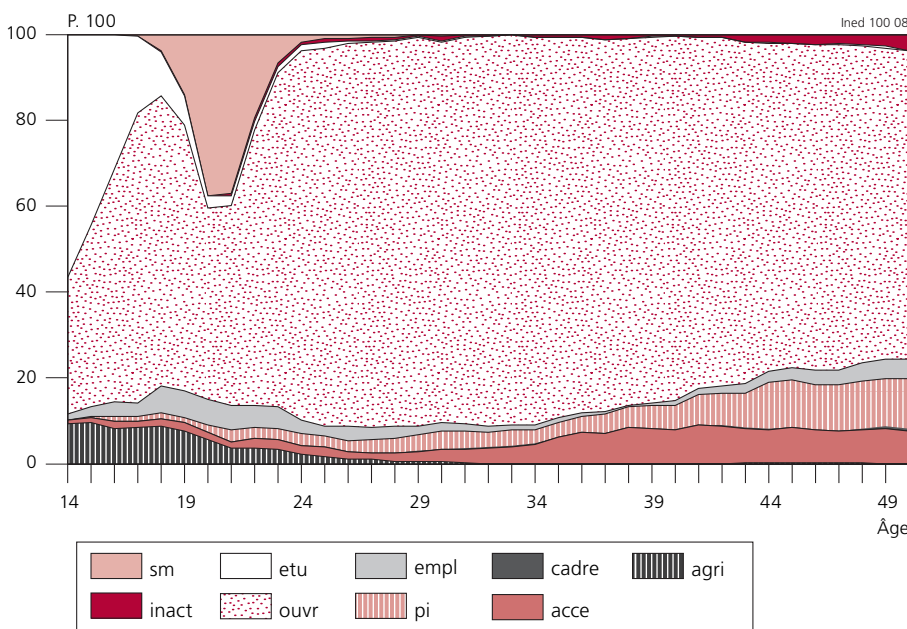
	agri	acce	cadre	pi	empl	ouvr	serv mil	inact	etu
agri	0,000	1,992	2,000	2,000	1,990	1,895	1,999	2,000	1,954
acce	1,992	0,000	1,990	1,986	1,987	1,978	1,997	1,987	1,991
cadre	2,000	1,990	0,000	1,971	1,990	1,997	1,972	1,964	1,912
pi	2,000	1,986	1,971	0,000	1,961	1,976	1,966	1,948	1,853
empl	1,990	1,987	1,990	1,961	0,000	1,970	1,972	1,962	1,896
ouvr	1,895	1,978	1,997	1,976	1,970	0,000	1,947	1,905	1,782
serv mil	1,999	1,997	1,972	1,966	1,972	1,947	0,000	1,986	1,947
inact	2,000	1,987	1,964	1,948	1,962	1,905	1,986	0,000	1,980
etu	1,954	1,991	1,912	1,853	1,896	1,782	1,947	1,980	0,000
* <i>indel</i> = 0,2 Lecture : agri = agriculteur ; acce = artisan, commerçant ou chef d'entreprise ; pi = profession intermédiaire ; empl = employé ; ouvr = ouvrier ; serv mil = service militaire ; inact = inactivité ; etu = étudiant. Champ : 1 341 carrières professionnelles d'hommes des générations 1930-1950 résidant en Île-de-France à la date de l'enquête. Source : enquête <i>Biographies et entourage</i> (Ined, 2001).									

Annexe 2. Chronogrammes de la typologie en 6 classes obtenue avec la MAO (proportion d'individus d'une classe donnée appartenant à chaque catégorie selon l'âge)

Classe 1



Classe 2



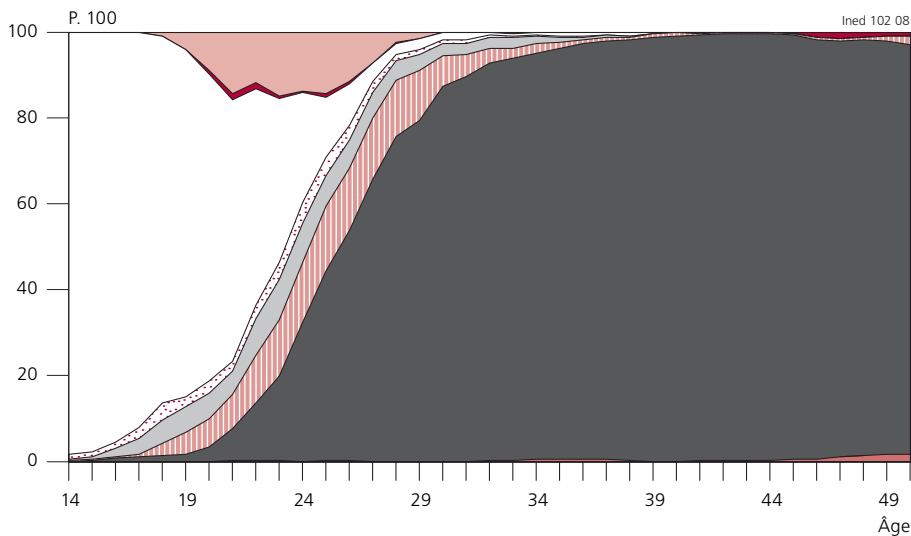
Lecture : sm = service militaire ; inact = inactivité ; etu = étudiant ; ouvr = ouvrier ; empl = employé ; pi = profession intermédiaire ; acce = artisan, commerçant ou chef d'entreprise ; agri = agriculteur.

Champ : hommes des générations 1930-1950 résidant en Île-de-France à la date de l'enquête.

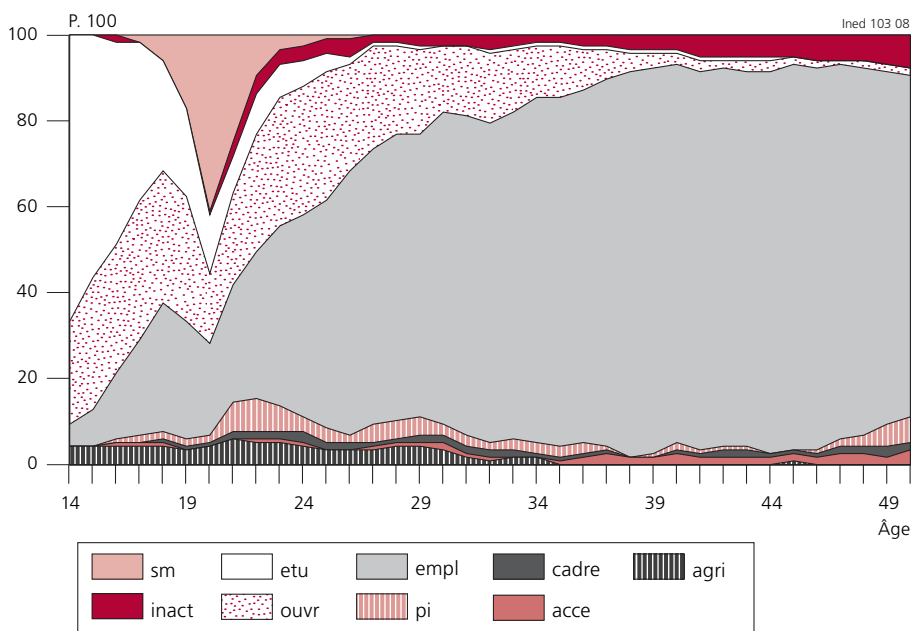
Source : enquête *Biographies et entourage* (Ined, 2001).

**Annexe 2 (suite). Chronogrammes de la typologie en 6 classes
obtenue avec la MAO (proportion d'individus d'une classe
donnée appartenant à chaque catégorie selon l'âge)**

Classe 3



Classe 4



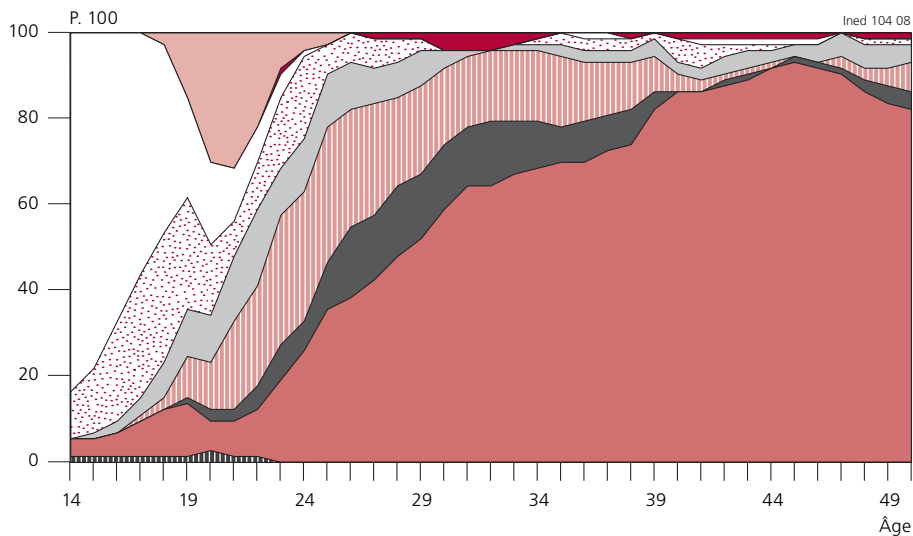
Lecture : sm = service militaire ; inact = inactivité ; etu = étudiant ; ouvr = ouvrier ; empl = employé ; pi = profession intermédiaire ; acce = artisan, commerçant ou chef d'entreprise ; agri = agriculteur.

Champ : hommes des générations 1930-1950 résidant en Île-de-France à la date de l'enquête.

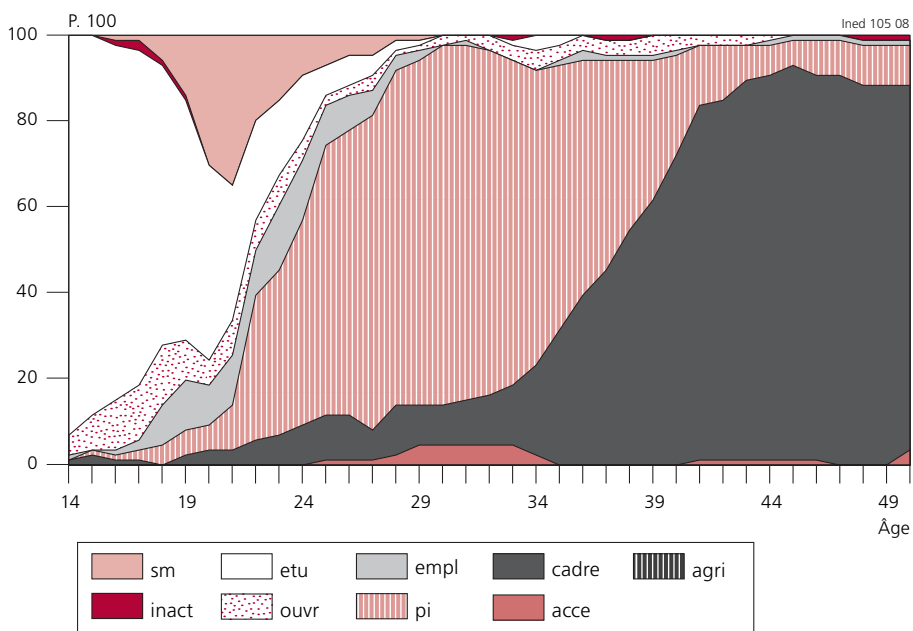
Source : enquête *Biographies et entourage* (Ined, 2001).

**Annexe 2 (suite). Chronogrammes de la typologie en 6 classes
obtenue avec la MAO (proportion d'individus d'une classe
donnée appartenant à chaque catégorie selon l'âge)**

Classe 5



Classe 6



Lecture : sm = service militaire ; inact = inactivité ; etu = étudiant ; ouvr = ouvrier ; empl = employé ; pi = profession intermédiaire ; acce = artisan, commerçant ou chef d'entreprise ; agri = agriculteur.

Champ : hommes des générations 1930-1950 résidant en Île-de-France à la date de l'enquête.

Source : enquête *Biographies et entourage* (Ined, 2001).

Annexe 3. Matrice de correspondance entre les typologies
en 10 classes obtenues par appariement optimal
et par analyse harmonique qualitative

		AHQ										
		1	2	3	4	5	6	7	8	9	10	Total
MAO	1	304	17	0	18	9	0	0	0	3	0	351
	2	4	187	0	0	15	0	0	0	2	0	208
	3	0	0	231	1	0	0	20	5	1	0	258
	4	1	2	0	83	0	0	0	0	1	0	87
	5	0	8	0	0	72	1	1	1	0	1	84
	6	1	10	0	0	0	41	0	0	0	21	73
	11	22	51	0	2	9	2	0	0	0	0	86
	12	0	6	27	4	24	24	3	0	0	6	94
	13	0	18	0	45	6	0	0	0	0	1	70
	14	0	0	5	13	3	0	2	4	3	0	30
Total		332	299	263	166	138	68	26	10	10	29	1 341
<i>Champ</i> : 1 341 carrières professionnelles d'hommes des générations 1930-1950 résidant en Île-de-France à la date de l'enquête. <i>Source</i> : enquête <i>Biographies et entourage</i> (Ined, 2001).												



RÉFÉRENCES

- AALLEN O., 1978, « Nonparametric inference for a family of counting processes », *The Annals of Statistics*, 6(4), p. 701-726.
- ABBOTT A., FORREST J., 1986, « Optimal matching methods for historical sequences », *Journal of Interdisciplinary History*, 16(3), p. 471-494.
- ABBOTT A., HRYCAK A., 1990, « Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers », *American Journal of Sociology*, (96), p. 144-185.
- BARBARY O., 1996, *Analisis tipologico de datos biograficos en Bogota*, Bogota, Universidad Nacional de Colombia, 254 p.
- BARBARY O., 1997, « Analisis estadistico de datos biograficos : metodos, ejemplos y perspectivas en el estudio de itinerarios migratorios » in J. A. Bustamante, D. Delaunay, J. Santibanez, *Medicion de la migracion internacional*, Tijuana, Documento de trabajo del Colegio de la Frontera Norte.
- BARBARY O., PINZON SARMIENTO L.M., 1998, « L'analyse harmonique qualitative et son application à la typologie des trajectoires individuelles », *Mathématiques informatiques et sciences humaines*, n° 144, p. 29-54.
- BILLARI F., 2001, « Sequence analysis in demographic research », *Canadian Studies in Population*, Special Issue on Longitudinal Methodology, 28(2), p. 439-458.
- BLAIR-LOY M., 1999, « Career patterns of executive women in finance : An optimal matching analysis », *American Journal of Sociology*, 104(5), p. 1346-1397.
- BONVALET C., BRINGÉ A., ROBETTE N., 2009, « Les trajectoires géographiques des Franciliens : un exemple de complémentarité qualitatif-quantitatif », in Actes du colloque *Approches quantitatives et qualitatives des mobilités*, Namur, AISLF, à paraître.
- CAMBOIS M.A., LELIÈVRE É., 1988, « Durée d'activité et interruption de carrière des femmes âgées de 45 ans à 69 ans en 1981 », *Population*, 43(3), p. 669-675.
- COURGEAU D., 2000, « Vers une analyse biographique multiniveau », in Actes des *Journées de méthodologie statistique*, Insee, 4-5 décembre, http://jms.insee.fr/site/files/documents/2008/658_1-JMS2000_S4-5-COURGEAU.PDF
- COURGEAU D., LELIÈVRE É., 1996, « Changement de paradigme en démographie », *Population*, 51(3), p. 645-654.
- COX D. R., 1972, « Regression models and life tables (with discussion) », *Journal of Royal Statistical Society*, B34, p. 187-220.
- CRÉPON B., GURGAND M., DEJEMEPPE M., 2005, « Counseling the unemployed : Does it lower unemployment duration and recurrence ? », *Document de travail*, Centre d'études de l'emploi, n° 40.
- DEGENNE A., LEBEAUX M.-O., 1999, *Étude sur les sorties du chômage, comparaison jeunes et adultes*, Rapport pour le Commissariat général du plan, Caen, Lasmas.
- DEGENNE A., LEBEAUX M.-O., MOUNIER L., 1996, « Typologies d'itinéraires comme instrument d'analyse du marché du travail », in A. Degenne, M. Mansuy, G. Podevin, P. Werquin (éd.), *Typologie des marchés du travail, suivi et parcours*, Document du Céreq n° 115, p. 27-42.
- DESPLANQUES G., SABOULIN M. (DE), 1986, « Activité féminine : carrières continues et discontinues », *Économie et Statistique*, n° 192-193, p. 51-62.

- DEVILLE J.-C., 1974, « Méthodes statistiques et numériques de l'analyse harmonique », *Annales de l'Insee*, n° 15, p. 3-101.
- DEVILLE J.-C., 1977, « Analyse harmonique du calendrier de constitution des familles en France », *Population*, 32(1), p. 17-63.
- DEVILLE J.-C., 1982, « Analyse de données chronologiques qualitatives : comment analyser des calendriers ? », *Annales de l'Insee*, n° 45, p. 45-104.
- DEVILLE J.-C., SAPORTA G., 1980, « Analyse harmonique qualitative », in E. Diday (éd.), *Data Analysis and Informatics*, Amsterdam, North Holland Publishing, p. 375-389.
- DIJKSTRA W., TARIS T., 1995, « Measuring the agreement between sequences », *Sociological Methods & Research*, (24), p. 214-231.
- DUREAU F., BARBARY O., ELISA FLOREZ C., HOYOS M. C., 1994, *La observacion de las diferentes formas de movilidad : propuestas metodologicas experimentadas en la encuesta de movilidad espacial en el area metropolitana de Bogota*, Paris, Orstom, Atelier du CEDE (Montevideo) du 27-29 octobre 1993 : « Nuevas modalidades y tendencias de la migracion entre paises fronterizos y los procesos de integracion », 31 p.
- ESPINASSE J.-M., 1993, « Enquêtes de cheminement, chronogrammes et classification automatique », *Note du Lhire*, 19(159).
- FLORETTE A., 1988, *Approximation et choix du découpage dans le cadre de l'analyse harmonique qualitative*, Mémoire de DEA, Ensae, Paris.
- GOUX D., 1991, « Coup de frein sur les carrières », *Économie et Statistique*, n° 249, p. 75-87.
- GRAB, 1999, *Biographies d'enquête. Bilan de quatorze collectes biographiques*, Paris, Ined (Méthodes et Savoirs, n° 3), 340 p.
- GRAB, 2006, *États flous et trajectoires complexes. Observation, modélisation, interprétation*, Paris, Ined (Méthodes et Savoirs, n° 5), 301 p.
- GRELET Y., 2002, « Des typologies de parcours. Méthodes et usages », *Document Génération* 92, (20), 47 p.
- HALPIN B., CHAN T. W., 1998, « Class careers as sequences : An optimal matching analysis of work-life histories », *European Sociological Review*, 14(2), p. 111-130.
- KAPLAN E., MEIER P., 1958, « Nonparametric estimation from incomplete observations », *Journal of American Statistical Association*, vol. 53, p. 457-481.
- KEMPENEERS M., LELIÈVRE É., 1991, « Analyse biographique du travail féminin », *Revue européenne de démographie*, 7, p. 377-400.
- LELIÈVRE É., 1987, « Activité professionnelle et fécondité : les choix et les déterminations chez les femmes françaises, de 1930 à 1960 », *Cahiers québécois de démographie*, 16, p. 209-236.
- LELIÈVRE É., VIVIER G., 2001, « Évaluation d'une collecte à la croisée du quantitatif et du qualitatif. L'enquête Biographies et entourage », *Population*, 56(6), p. 1043-1074.
- LESNARD L., SAINT-POL T. (DE), 2004, « Introduction aux méthodes d'appariement optimal (Optimal Matching Analysis) », *Document de travail du Crest*, (15), 30 p.
- MACINDOE H., ABBOTT A., 2004, « Sequence analysis and optimal matching techniques for social science data », in Hardy Melissa, Bryman Alan, *Handbook of Data Analysis*, London, Sage, p. 387-406.
- MARCHAND O., THÉLOT C., 1997, *Le travail en France. 1800-2000*, Paris, Nathan, 269 p.
- NELSON W., 1972, « Theory and application of hazard plotting for censored failure data », *Technometrics*, vol. 14, p. 945-965.
- ROHWER G., PÖTTER U., 2005, « TDA's user manual », <http://www.stat.ruhr-uni-bochum.de/tman.html>

- SANKOFF D., KRUSKAL J. (DIR.), 1983, *Time Warps, String Edits and Macromolecules : The Theory and Practice of Sequence Comparison*, Reading, Addison-Wesley, 408 p.
- SCHERER S., 2001, « Early career patterns : A comparison of Great Britain and West Germany », *European Sociological Review*, 17(2), p. 119-144.
- SOLIS P., BILLARI F., 2002, « Work lives amid social change and continuity: Occupational trajectories in Monterrey, Mexico », *Max Planck IDR Working paper*, 2002-009, 52 p.
- STOVEL K., SAVAGE M., BEARMAN P., 1996, « Ascription into achievement : Models of career systems at Lloyds Bank, 1890-1970 », *American Journal of Sociology*, 102(2), p. 358-399.

NICOLAS ROBETTE, NICOLAS THIBAUT • ANALYSE HARMONIQUE QUALITATIVE OU MÉTHODES D'APPARIEMENT OPTIMAL ? UNE ANALYSE EXPLORATOIRE DE TRAJECTOIRES PROFESSIONNELLES

Les enquêtes biographiques permettent d'analyser un grand nombre de carrières professionnelles individuelles dans leur intégralité. Diverses méthodes statistiques ont été développées pour mesurer les durées de séjour dans un état considéré en fonction de caractéristiques individuelles. Jusqu'aux années 1990, le traitement exploratoire des données avec l'objectif de décrire les parcours dans leur complexité n'avait fait l'objet que de peu d'attention dans la littérature. L'analyse harmonique qualitative et les méthodes d'appariement optimal sont deux méthodes exploratoires qui permettent de construire des typologies de parcours individuels complexes en prenant en compte la séquence des événements et leur durée. On les utilise ici pour reconstituer des typologies des carrières professionnelles des hommes interrogés lors de l'enquête *Biographies et entourage* (Ined, 2001) afin de comparer l'intérêt respectif de chacune de ces techniques.

NICOLAS ROBETTE, NICOLAS THIBAUT • COMPARING QUALITATIVE HARMONIC ANALYSIS AND OPTIMAL MATCHING. AN EXPLORATORY STUDY OF OCCUPATIONAL TRAJECTORIES

Event history surveys provide a means to analyse large numbers of complete individual occupational trajectories. A variety of statistical methods have been developed to measure the time spent in a give state as a function of individual characteristics. Until the 1990s, exploratory data analysis to describe the full complexity of trajectories was rarely mentioned in the literature. Qualitative harmonic analysis and optimal matching are two exploratory methods that can be used to build typologies of complex individual trajectories that take account of both the sequence and the duration of events. They are used here to classify the working careers of male respondents of the *Biographies et entourage* survey (INED, 2001), with the aim of comparing the respective advantages of each technique.

NICOLAS ROBETTE, NICOLAS THIBAUT • ¿ ANÁLISIS ARMÓNICO CUALITATIVO O MÉTODOS DE EMPAREJAMIENTO ÓPTIMO ? UN ANÁLISIS EXPLORATORIO DE TRAYECTORIAS PROFESIONALES

Las encuestas biográficas permiten analizar un gran número de carreras profesionales individuales en su totalidad. Se han desarrollado diversos métodos estadísticos para medir las duraciones de las estancia en un estado considerado en función de características individuales. Hasta los años 1990, el tratamiento exploratorio de los datos con el objetivo de describir los recorridos en su complejidad había sido objeto de poca atención en la literatura. El análisis armónico cualitativo y los métodos de emparejamiento óptimo son dos métodos exploratorios que permiten elaborar tipologías de recorridos individuales complejos tomando en cuenta la secuencia de los acontecimientos y su duración. Aquí se les utiliza para reconstituir tipologías de las carreras profesionales de los hombres interrogados en la encuesta *Biographies et entourage* (Ined, 2001), con la finalidad de comparar el interés respectivo de cada una de estas técnicas