

DECRIRE DES DONNEES SEQUENTIELLES EN SCIENCES SOCIALES : MISE EN PRATIQUE DES METHODES D'APPARIEMENT OPTIMAL

Thibaut DE SAINT POL (), Laurent LESNARD (**)*

() Insee, Conditions de vie des ménages
Crest, Laboratoire de sociologie quantitative
(**) Sciences Po, Centre de données sociopolitiques
Crest, Laboratoire de sociologie quantitative*

Introduction

Le développement dans les pays anglo-saxons de techniques pour décrire les données séquentielles ouvre de nouvelles possibilités pour le statisticien et le chercheur. Ainsi, l'utilisation en sciences sociales d'une technique auparavant utilisée en informatique et en biologie moléculaire, les méthodes d'appariement optimal (M.A.O.)¹ (en anglais *Optimal matching analysis*), pose la question de son application pratique aux données propres à ce domaine.

Reposant sur un principe simple et l'automatisation des opérations que l'on fait intuitivement pour comparer des séquences entre elles, les méthodes d'appariement d'optimal sont pourtant encore assez peu utilisées en France. Au travers d'exemples, cet article a pour objectif de compléter la présentation² plus théorique réalisée lors des Journées de méthodologie statistique d'avril 2009 en montrant comment la flexibilité de cette méthode permet de l'adapter avec pertinence à des données et des questions très diverses.

Les méthodes d'appariement optimal peuvent en effet s'appliquer à de nombreux objets dès lors que l'étude porte sur une succession d'états ou d'actions. Ce peut être le cas à l'échelle d'une vie ou d'une portion de vie. On peut par exemple étudier les différents métiers qu'ont exercés des individus ou les différents lieux dans lesquels ils ont habité. Mais on peut également raisonner sur des périodes plus courtes : une année, un mois ou une semaine.

Les deux applications qui seront plus particulièrement développées dans le cadre de cette communication portent sur une durée encore plus courte : elles s'intéressent aux activités réalisées pendant une journée à partir des deux dernières enquêtes Emploi du Temps de l'Insee. La première portera sur la description des journées de travail en France. Dans le second, nous nous intéresserons à l'inscription des pratiques alimentaires dans la séquence des activités de la soirée des Français. D'autres applications aux carrières et aux données biographiques seront ensuite également présentées.

¹ en anglais *Optimal matching analysis* (O.M.A.).

² Laurent LESNARD, Thibaut DE SAINT POL, *Décrire des données séquentielles en sciences sociales : panorama des méthodes existantes*, Session « Analyses longitudinales ».

1. Le cadre d'étude

1.1. Deux applications sur les mêmes données

Les deux applications qui seront présentées dans ce papier s'intéressent aux activités réalisées pendant une journée à partir des deux dernières enquêtes Emploi du Temps de l'INSEE. L'emploi du temps est en effet un bon exemple de matériel séquentiel : tout au long de la journée, les individus enchaînent des activités plus ou moins longues. Calculer les moyennes des temps consacrés à chaque activité, comme on le fait fréquemment, ne suffit cependant pas à rendre compte de la diversité des organisations quotidiennes. Travailler huit heures en continu n'est pas pareil que de consacrer quatre plages de deux heures au travail à divers moments de la journée.

Les méthodes d'appariement optimal permettent justement de tirer parti de la spécificité de telles données. Les études développées ci-dessous proposent deux utilisations différentes de cette méthode à partir du même matériau. La présentation de chacun de ces exemples mériterait lui-même beaucoup plus d'espace : on se concentrera ici sur les apports empiriques des M.A.O. et l'on ne fera qu'esquisser ~~ici~~ les éléments théoriques. Le premier exemple peut être considéré comme une application typique des techniques d'appariement optimal : les deux ensembles d'opérations sont ici utilisés de manière à faire émerger la manière dont le dîner s'inscrit dans la soirée des Français. Le second, au contraire, illustre la souplesse de la technique d'appariement de séquences et propose d'identifier les différents types d'horaires de travail à l'aide des seules opérations de substitutions, de manière à préserver au maximum leurs dimensions temporelles.

1.2. La mise en œuvre informatique

C'est dans le cadre de la bio-informatique que la plupart des logiciels d'appariement de séquences ont été développés et continuent de l'être. La vigueur de la recherche appliquée dans cette discipline est telle qu'il est impossible de proposer un inventaire exhaustif de logiciels pour la plupart très spécialisés. Citons un des programmes les plus anciens *Clustal*, qui semble avoir suivi le rythme vertigineux des améliorations de ce champ, et un plus récent, *EMBOSS* (European Molecular Biology Open Software Suite). Promoteur de ces méthodes en sciences sociales, Andrew Abbott a supervisé le développement d'un logiciel, *Optimize*, qui n'a cependant plus évolué depuis 1997. Le logiciel de statistique libre *TDA* proposé par Goetz Rohwer et Ulrich Pötter de l'université de Bochum comporte un module d'analyse de séquence pourvu de quelques fonctionnalités d'appariement optimal. C'est ce logiciel qui a été utilisé pour le premier exemple, complété par le logiciel SAS pour les manipulations des données et la classification des distances issues de TDA. Le second exemple a été entièrement mené à bien sous SAS, les simplifications supplémentaires facilitant encore davantage la mise en œuvre informatique de l'appariement.

D'autres logiciels peuvent aujourd'hui être utilisés pour mettre en œuvre les MAO. L'extension (*plugin*) *sq* apporte les principales variantes des MAO au logiciel *Stata* [1]. La relative lenteur de *sq* peut être améliorée en utilisant les extensions écrites en C par Brendan Halpin [2]. La variante des MAO appelée *Dynamic Hamming Matching* est disponible pour *Stata* sous forme d'extension [3] et dans la librairie *TraMineR* du logiciel libre *R* [4]. Cette librairie fait de *R* le logiciel le plus adaptée à la mise en œuvre des MAO. Les nombreuses variantes de MAO disponibles sont en outre très rapides. *TraMineR* offre également de nombreuses autres procédures pour décrire et représenter graphiquement les séquences.

2. L'étude du repas dans le cadre de la soirée

L'étude des comportements alimentaires s'effectue généralement indépendamment de l'analyse des autres temps quotidiens. On analyse fréquemment les durées moyennes que les femmes dédient à l'alimentation par rapport à celles des hommes, ou encore celles des plus jeunes par rapport aux plus vieux, sans s'intéresser aux activités qui encadrent les repas. Or il existe une forte interdépendance en termes d'horaires, de lieu, ou même de compagnie entre les différentes activités qui composent notre emploi du temps. Si le repas du midi s'effectue souvent à l'extérieur du domicile et dure moins longtemps en moyenne que le dîner, c'est parce qu'il s'inscrit généralement au milieu d'activités professionnelles qui influent directement sur la manière dont va se dérouler cette prise alimentaire. Le repas du soir quant à lui se déroule entre des activités plus diversifiées qui vont des travaux domestiques aux loisirs les plus divers, en passant par le travail professionnel ou même le sommeil.

Comprendre les logiques qui président à la réalisation du dîner, ce n'est pas seulement étudier sa durée, l'heure de son commencement ou encore calculer le temps moyen que lui consacre telle ou telle catégorie de la population. On ne peut également se contenter d'une analyse du type que celles que permettent les régressions. Expliquer les pratiques alimentaires en fonction de paramètres comme l'âge, le sexe ou encore le lieu de résidence comporte un intérêt certain. Mais ces démarches négligent la possibilité d'une causalité multiple inscrite dans la temporalité : le fait que tel individu dîne sur telle plage horaire est le résultat d'un processus auquel ont contribué, par exemple, le fait qu'il ait quitté tard le bureau, qu'il soit ensuite resté une heure dans les embouteillages, mais qu'il ne veuille en aucun cas manquer le film qui débute à 20h50. Ce sont ces contraintes imposées par les activités qui encadrent le repas qui permettent de comprendre la manière dont se déroule le dîner et qui participent par leur récurrence à la construction d'habitudes alimentaires. Une régression ferait par exemple apparaître la corrélation entre les repas les plus courts et le fait d'être actif. Mais elle oublierait que l'explication de cette relation se situe dans l'enchaînement des différentes activités.

Afin de décrire le contexte dans lequel s'effectue le dîner, cette étude s'est intéressée à la période 18h50-21h30 pendant laquelle se concentrent les prises alimentaires de la soirée [6]. Les activités présentes dans les carnets journaliers de l'enquête Emploi du Temps 1998 ont été regroupées en 25 catégories³ qui sont autant d'éléments possibles constitutifs de nos séquences. L'appariement optimal porte sur des séquences de même taille, c'est-à-dire de 16 éléments qui correspondent aux 16 plages horaires de dix minutes de la période étudiée. À aucun moment de ce traitement séquentiel, les plages alimentaires ne seront privilégiées dans le regroupement des séquences. Ce point qui pourrait paraître anodin est en fait fondamental. Trop de classifications font apparaître des dissemblances qui découlent directement de ce choix d'agrégation et qui biaisent l'analyse sociologique qui s'y rapporte. Ici, le regard du sociologue n'intervient pas dans le processus de regroupement des séquences.

Mais les M.A.O. ont un autre intérêt : c'est une technique particulièrement flexible qui s'adapte très aisément aux contraintes imposées par le matériau utilisé et à la théorie au travers du choix des coûts des différentes opérations. Ici, nous avons calculé les coûts de chacune des trois opérations en termes de fréquences des différents éléments constitutifs des séquences. Ainsi, le coût de substitution d'un épisode de travail par un épisode de sommeil, situation peu courante dans nos emplois du temps, sera élevé. Au contraire, le coût de la substitution d'un repas par un épisode de télévision sera plus faible⁴. Ces coûts sont donc calculés sur l'ensemble des 16 épisodes : ils ne tiennent pas compte de la tranche horaire considérée. Il pourrait en effet être plus probable que la télévision suive le repas à 21h00 qu'à 19h00. C'est pourquoi nous l'avons introduite dans les coûts d'insertion et de suppression. Il est important pour notre étude des rythmes que l'algorithme puisse prendre en compte cette dimension temporelle. Dans les emplois du temps, une opération n'a pas toujours le même coût quelle que soit l'heure à laquelle l'activité se déroule.

³ Ces catégories sont celles proposées par Alain Chenu et Nicolas Herpin [5].

⁴ En termes mathématiques, ces coûts sont les inverses des probabilités de transition entre deux activités sur l'ensemble des séquences de notre échantillon.

Or ce sont les insertions et les suppressions qui, en décalant les activités de dix minutes à chaque fois, posent le problème de différence des tranches horaires. L'emploi de ces deux opérations est ce qui fait la singularité et l'originalité de cette analyse séquentielle. Mais une trop grande utilisation de ces mouvements revient à décaler totalement les éléments de nos séquences et à perdre de ce fait les particularités de chaque tranche horaire. Il nous fallait donc autoriser le recours à ces deux opérations, tout en empêchant une utilisation abusive. Nous avons choisi de rehausser légèrement les coûts d'insertion et de suppression par rapport aux substitutions pour que ces dernières soient privilégiées, suivant en cela les recommandations d'Abbott et Hrycak [7].

L'étude comparative des résultats avec ou sans cette modification montre qu'elle augmente la variance inter-classe et diminue celle intra-classe : limiter le recours à l'insertion et à la suppression a permis d'éviter des classements abusifs et a amélioré l'homogénéité de nos classes. Ce jeu sur le coût des opérations est moins complexe qu'il ne peut paraître de prime abord. C'est un des intérêts de cette méthode d'analyse. Si l'objet d'étude n'intervient pas directement dans la construction des classes, il est néanmoins possible d'adapter simplement l'algorithme de la méthode aux particularités temporelles de cet objet afin de coller au plus près à la réalité sociale que traduit la séquence.

Ces choix ont conduit à regrouper dans notre exemple les séquences en dix classes que l'on peut résumer très imparfaitement à partir du tableau 1⁵. Mis à part les individus de la première classe qui consacrent beaucoup de temps au repas et les deux dernières où le temps alimentaire est faible, le recours par exemple aux moyennes voile un certain nombre de pratiques fort diverses que le recours à une analyse séquentielle permet de mettre en lumière. Ainsi, les temps moyens consacrés au repas pour les quatrième et huitième classes sont très proches. Pour autant, le repas est pris en début de période pour le premier groupe et en seconde partie pour le deuxième groupe. Cette observation simple suffit à éclairer les possibilités nouvelles qu'offrent les M.A.O. Ainsi, au sein de ce qui n'était auparavant qu'un groupe assez difforme d'individus aux temps moyens alimentaires similaires, il est désormais possible de distinguer différents types de pratiques.

Classe	Effectif (en %)	Temps moyen consacré au REPAS de 18h50 à 21h30	Nom de la classe	Séquence-type ⁶				
				19h00- 19h30	19h30- 20h00	20h00- 20h30	20h30- 21h00	21h00- 21h30
1	6,0	113 min	<i>Les Mangeurs</i>	Repas	Repas	Repas	Repas	Repas
2	3,7	43 min	<i>Les Couche-tôt</i>	Repas	Repas	Télévision	Sommeil	Sommeil
3	20,8	37 min	<i>Soirée télé</i>	Repas	Repas	Télévision	Télévision	Télévision
4	12,0	42 min	<i>Les Dîne-tôt</i>	Repas	Repas	Télévision	Télévision	Télévision
5	5,8	48 min	<i>Cuisine et Ménage</i>	Repas	Repas	Ménager	Enfants	Télévision
6	24,5	36 min	<i>Les Dîne-tard</i>	Bricolage	Repas	Repas	Repas	Télévision
7	7,5	40 min	<i>Deuxième journée</i>	Ménager	Ménager	Repas	Repas	Télévision
8	5,4	43 min	<i>Les Téléphages</i>	Télévision	Télévision	Repas	Repas	Télévision
9	9,4	26 min	<i>Les Travailleurs</i>	Travail	Travail	Travail	Travail	Travail
10	4,9	29 min	<i>Sorties</i>	Rencontres	Rencontres	Rencontres	Rencontres	Rencontres

Source : Enquêtes Emploi du Temps de l'Insee de 1998.

Tableau 1 – Descriptif des dix classes.

⁵ Pour plus de précision, voir [6].

⁶ La séquence-type reproduite ici est celle qui se situe au centre de la classe : elle minimise la distance à toutes les autres séquences du groupe. Dans une visée illustrative, on a enlevé le premier élément qui correspond à la plage 18h50-19h00 et qui est identique à l'élément suivant pour toutes les séquences, mis à part pour la 3^e et la 4^e où il s'agit de travaux ménagers, et on a regroupé les éléments trois par trois.

Chacune de ces soirées-type fait apparaître une logique d'insertion du dîner dans la soirée et montre l'importance du contexte dans lequel se déroule cette prise alimentaire. Cette approche nouvelle est d'autant plus intéressante que l'on peut caractériser ces séquences au moyen des caractéristiques des individus auxquelles elles appartiennent. On va ainsi pouvoir opposer par exemple des séquences d'activités féminines, marquées par le poids du travail ménager comme la classe 7 par exemple, s'opposant à des séquences plus masculines comme la classe 8, où la télévision occupe une grande place.

Par ailleurs, le recours à une méthode d'appariement optimal permet de dépasser certaines limites propres aux analyses classiques. Dans le cas de notre exemple, le passage entre l'enquête Emploi du Temps de 1985 et celle de 1998 de l'interligne du carnet journalier rempli par les personnes interrogées de cinq à dix minutes conduit à un biais méthodologique qui empêche de mener des comparaisons satisfaisantes des durées des activités entre les deux enquêtes. En effet, les activités les plus courtes, comme mettre ou débarrasser la table ou plus généralement le travail domestique qui est très fractionné, ont été souvent intégrées par les enquêtés dans des activités plus longues. Ainsi, on observe en moyenne entre 1985 et 1998 une augmentation de près de dix minutes du temps consacré aux repas, qui est plus que suspecte⁷. L'utilisation de M.A.O. rend possible le dépassement de ce biais méthodologique en ne considérant pour 1985 qu'une ligne pour deux ; ce qui a pour conséquence directe de faire disparaître la moitié des activités de cinq minutes et de prendre en compte leur relative disparition en 1998. L'application du même protocole réalisé pour les soirées des Français en 1998 pour les données de 1985 amène ainsi à la construction d'une typologie extraordinairement proche de celle présentée ci-dessus. Ce qui offre d'ailleurs une possibilité de contrôle de la grande robustesse des classes obtenues au moyen de M.A.O.. Ce bref aperçu de l'approche novatrice autorisée par l'utilisation de cette technique pour la compréhension des pratiques alimentaires milite pour son adoption et sa mise en pratique à d'autres objets de recherche.

3. Le temps du travail

L'analyse scientifique du temps de travail est généralement réduite dans les enquêtes Emploi du temps à de simples durées⁸, ce qui occulte nombre de variations [9]. Ainsi, des positions *a priori* opposées comme celles de la diminution du temps consacré au travail [10] et de l'extension du *workaholism*⁹ [11] peuvent-elles être réconciliées dès lors que les moyennes nationales sont décomposées selon la position sociale ou le niveau d'éducation : la thèse du renversement du gradient du niveau d'éducation-travail [12,13] permet à cet effet de réconcilier ces deux théories en soulignant les changements des rapports entretenus entre position dans la hiérarchie sociale et temps de travail.

Toutefois, cette décomposition de moyenne ne permet pas de relier les évolutions des heures moyennes travaillées avec un autre thème majeur, celui de la *flexibilité*, des horaires de travail notamment. La moyenne ne permet donc pas d'appréhender le travail dans son déroulement, de connaître la répartition des heures travaillées dans la journée. De la même manière, les indicateurs de flexibilité apparaissent sensibles à la durée du travail et à la répartition du travail dans la journée : puisqu'ils sont construits *a priori*, les indicateurs de flexibilité sont bien souvent hétérogènes. On peut citer l'exemple du travail de nuit des Enquêtes Emploi : une personne travaille de nuit si sa période d'activité se situe, même partiellement, entre minuit et cinq heures du matin. Les journées de travail qui commencent à cinq heures (horaires décalés le matin), celles qui se terminent à minuit (horaires décalés le soir) se trouvent ainsi mélangées au véritable travail de nuit.

⁷ Il faut donc interpréter avec prudence les résultats de G. Larmet qui conclut, uniquement en termes de durée, à l'accroissement de la sociabilité alimentaire entre 1985 et 1998 [8].

⁸ La remarque s'applique également aux dernières enquêtes Emploi de l'Insee dont la question sur l'heure de début et de fin du travail se transforme invariablement dans les exploitations en simple durée.

⁹ Terme anglo-saxon qui désigne les travailleurs compulsifs.

Seule une classification peut conjuguer régularité statistique et diversité et dépasser ainsi l'antagonisme de la moyenne et des indicateurs *a priori*. Pour mesurer la dissimilarité des journées de travail en termes de durée mais également de répartition des heures travaillées dans la journée, l'approche séquentielle des méthodes d'appariement optimal semble idéale. Le recodage binaire (travail/non-travail) des carnets des enquêtes Emploi du temps de 1985 et 1998 associé à un algorithme d'appariement optimal devrait donc permettre de construire une typologie des horaires de travail en France.

Toutefois, comme il ne s'agit pas ici d'identifier des enchaînements typiques, les journées étant stylisées à l'extrême à l'opposé de l'exemple précédent, mais au contraire d'identifier les *décalages* temporels du travail, les opérations de substitutions doivent être fortement privilégiées au détriment des opérations d'insertion-suppression (qui brouilleraient les décalages des horaires de travail). Mieux, les opérations d'insertion-suppression peuvent être bannies du processus d'appariement puisque seule la dimension temporelle du travail nous intéresse ici. Cet exemple illustre donc particulièrement bien la souplesse de l'analyse d'appariement qui, dans ce cas très particulier, n'est plus *optimal*¹⁰.

Reste à déterminer les différents coûts de substitution entre les deux états travail et non-travail. Si la théorie sociologique ne semble pas ici en mesure de déterminer directement de tels coûts, elle peut cependant guider leur construction : puisque, comme l'a montré Durkheim [14], le temps est un système symbolique qui, parce qu'il cristallise le rythme de l'activité collective, permet d'anticiper les régularités sociales, c'est le rythme collectif qui va fournir le moyen de différencier les différents emplois du temps de travail. En effet, la traduction du postulat durkheimien de la différenciation sociale du temps, autrement dit la différenciation du flux incessant d'événements par l'activité collective, en des termes plus opérationnels nous donne un moyen de détermination des coûts de substitution : c'est la position relative des emplois du temps individuels par rapport au rythme collectif qui va nous donner une mesure de la similarité des emplois du temps.

Le rythme de l'activité collective qui nous intéresse ici est le rythme du travail et peut être approché simplement par les « flux » entre les deux états « travail » et « non-travail » : un flux élevé entre ces deux états signifie qu'un changement de rythme est en cours donc qu'un travailleur et un inactif sont assez proches puisqu'ils risquent de partager le même état¹¹. Au contraire, une faible circulation entre ces états est signe d'un certain hermétisme (les deux rythmes coexistent), ce qui fait qu'un travailleur et un inactif seront alors éloignés. Par exemple, la transition entre travail et non-travail a de bonnes chances d'être élevée vers 9h, ce qui va limiter la distance entre un travailleur et un inactif. En revanche, vers 3h ou 15h, cette même transition sera très vraisemblablement plus faible, ce qui accentuera la différence entre un travailleur et un inactif à de telles heures. L'appariement des horaires proposé s'accorde donc avec le sens commun qui voit la différence comme un écart à la norme : des horaires ne deviennent atypiques qu'en relation à une norme collective de rythme de travail. Plutôt que de la fixer arbitrairement, la norme émerge ici des régularités observées : c'est le rythme collectif qui va déterminer le degré de différence entre deux horaires de travail : la mesure de dissimilarité proposée est donc à la fois endogène et dynamique¹². Au final, la distance entre deux emplois du temps individuels est obtenue par la somme de leurs différences instantanées, i.e. par la suite de leurs positions relatives par rapport aux rythmes temporels du champ considéré.

La mise en œuvre qui vient d'être décrite, associée à l'algorithme WPGMA flexible de classification ascendante hiérarchique nous permet d'identifier douze horaires de travail typiques. Ces types peuvent être décrits à l'aide de deux indicateurs : la mi-journée de travail et la durée de cette journée de travail, autrement dit par un indicateur de position centrale et un autre indiquant la dispersion autour de cette tendance. L'interprétation de la plupart des douze classes est aisée (voir Tableau 2 *supra*¹³).

¹⁰ En effet, sans opérations d'insertion-suppression, un seul chemin est possible : celui situé sur la diagonale de la matrice d'appariement.

¹¹ En termes statistiques, ces flux sont mesurés par les matrices de transition entre les différents états. Pour plus de détails, voir [15,16].

¹² Cette version des M.A.O. peut être vue comme un cas particulier de la distance de Hamming pondérée par la série des matrices de transition entre épisodes. Pour une présentation plus complète et technique, voir [15,16]. Cette méthode est disponible sur Internet sous la forme d'une extension Stata (voir <http://laurent.lesnard.free.fr>) et dans la librairie *TraMineR* du logiciel R [4].

¹³ Seuls les résultats pour 1985-86 sont présentés ici. Pour plus de détails, voir [17,18]

No. classe	Type d'horaire de travail	Effectifs (% de la pop. tot.)	Mi-journée de travail	Durée de travail
	Standard	56,5%	12:59	8:26
1	7-16	7,6%	12:00	8:14
2	8-18	38,2%	12:53	8:17
3	9-19	10,7%	14:01	9:09
	Décalé	14,4%		7:16
4	Matin	5,3%	9:44	7:39
5	Après-midi	5,4%	15:32	6:46
6	Soir	2,1%	17:02	7:20
7	Nuit	1,7%		7:38
	Extensif	9,1%	13:57	10:29
8	Régulier	3,5%	12:54	10:47
9	Irrégulier	5,6%	14:38	10:18
	Irrégulier	20,0%	12:50	3:45
10	Fragmenté	3,2%	13:21	3:50
11	Étalé	3,5%	12:15	8:06
12	Faible durée	13,3%	12:52	2:14

Source : Enquêtes Emploi du Temps de l'Insee de 1985-86.

Lecture : la première classe (No. 1) appartient au sous-groupe des horaires standards et représente 7,6% des journées travaillées. La mi-journée de travail de ce type d'horaire se situe en moyenne à midi alors que sa durée moyenne est de huit heures et quart.

Tableau 3 – Principales caractéristiques des douze types d'horaires de travail.

Les trois premiers types constituent des horaires standard correspondant à une journée de travail de huit heures avec des horaires de bureau centrés autour de la mi-journée (13h) et regroupent un peu plus de la moitié des journées travaillées. Ces trois types d'horaires de travail apparaissent conformes à ce que l'on considère comme une « journée de travail normale » : la technique d'appariement optimal permet donc d'isoler les horaires de travail les plus courants, dont les caractéristiques apparaissent conformes à la norme tacite des horaires de travail « normaux ».

La déviance la plus conséquente à cette norme de journée de travail repose essentiellement sur une divergence temporelle considérable de la mi-journée de travail par rapport à la mi-journée « normale » qui se situe ici aux alentours de 13h. Ces types d'horaires peuvent être considérés comme atypiques et contiennent notamment le travail de nuit¹⁴ qui ne correspond pas du tout aux définitions classiques retenues usuellement. Dans les enquêtes Emploi de l'Insee, est considéré comme travail de nuit toute période de travail située, même partiellement, entre minuit et cinq heures du matin : l'analyse proposée ici permet en quelque sorte d'affiner cette catégorie avec laquelle elle se superpose en partie, mais surtout, parce qu'elle ne repose pas sur des règles strictes fixées *a priori*, elle augmente significativement le nombre des horaires atypiques¹⁵. La durée moyenne inférieure à huit heures indique que ces horaires de travail contiennent une proportion importante de journées partiellement travaillées, autrement dit que la réduction du temps de travail s'accompagne d'une marginalisation de la répartition de ces heures travaillées dans la journée.

Mais les horaires décalés ne sont pas la seule source de déviance par rapport aux horaires de travail « normaux » : les longues journées de travail peuvent également être légitimement considérées

¹⁴ Le travail de nuit est ici très particulier puisque deux « journées » de travail sont partiellement observées, le travail de nuit ne correspondant pas à la fenêtre d'une journée des enquêtes Emploi du Temps françaises.

¹⁵ À peu près 20 % des horaires non-standard identifiés par la classification entrent dans la définition du travail de nuit de l'enquête emploi. Par conséquent, identifier les horaires atypiques aux seules périodes de travail de nuit limite singulièrement l'appréciation de l'importance des horaires décalés. Si, par définition, les horaires de nuit sont complètement inclus dans le travail de nuit, seuls 10 % des horaires du matin et 30 % des horaires du soir entrent dans le champ du travail de nuit, sans parler de l'exclusion des horaires décalés dans l'après-midi.

comme atypiques. Deux classes d'horaires de travail présentent ainsi une durée de travail supérieure à dix heures, situation qui représente près de 10 % des journées travaillées. De même, les petites journées de travail apparaissent non-standard, anomalie de durée parfois redoublée par une fragmentation de ce travail au cours de la journée. Ces types d'horaires sont la conséquence du processus de sélection des journées travaillées (une journée est considérée comme travaillée dès lors qu'elle présente au moins une déclaration de travail dans le carnet d'emploi du temps) et contiennent un nombre non négligeable de séances de travail le week-end de cadres ou d'enseignants de même que d'horaires de travail fragmentés de certaines catégories d'employés comme les caissières qui peuvent enchaîner deux séances de travail 10-13h et 16-20h dans une journée [19,20].

Ainsi, contrairement à l'image véhiculée par les indicateurs construits à partir de règles rigides, les horaires atypiques, loin d'être minoritaires, représentent une part presque équivalente à la journée de travail « normale ». Parce qu'ils sont trop synthétiques, indicateurs et moyennes fragmentent et figent le travail, autrement dit offrent une vision partielle des transformations du travail. Seule une approche en termes de séquence permet de lier les changements de la durée de la journée de travail avec la flexibilité des horaires et d'apercevoir ainsi que la réduction de la durée de travail s'accompagne souvent d'une répartition non-standard de ces horaires.

Conclusion

Les méthodes d'appariement optimal, en mettant au premier plan la séquence au sein de l'analyse sociologique, permettent non seulement de décrire autrement les phénomènes sociaux, mais remettent en lumière la dimension temporelle de la causalité. Penser en séquences, c'est saisir l'action au travers de sa durée et de ses bornes, comme on le fait habituellement lorsque par exemple on calcule des moyennes ou que l'on fait des régressions. Mais c'est aussi considérer l'action parmi un enchaînement d'autres actions qui ont elles aussi une durée et des bornes qui influent sur les éléments qui les précèdent ou qui les suivent.

Une action ne se construit pas isolément dans une boîte noire en fonction de facteurs tels que l'âge, le sexe ou la profession du sujet. Elle est presque toujours déterminée par la suite d'actes dans laquelle elle s'inscrit. Ainsi, c'est parce qu'elle n'a pas un revenu suffisant que telle employée doit se passer des services d'une nourrice et qu'elle doit chaque matin conduire ses enfants à l'école. Et c'est parce que l'école de ses enfants ouvre à un horaire fixe que cette personne arrive fréquemment en retard au travail. Le processus de causalité s'établit dans la chronologie. C'est précisément cette chronologie que les méthodes d'appariement optimal permettent de mettre en lumière.

L'intérêt de cette technique ne se limite pas à l'étude des emplois du temps. Elles s'appliquent à toutes les données dynamiques, notamment aux carrières [7, 21]. Si le principe de ces méthodes repose sur l'optimisation des opérations élémentaires engagées dans toute comparaison manuelle de séquences – insertion, suppression et substitution – l'automatisation de ce traitement exige que soient explicitées les règles de la comparaison au travers des coûts qui sont affectés à ces opérations. À cet égard, les méthodes d'appariement optimal permettent de réconcilier les oppositions artificielles entre théorie et pratique, et entre traitement quantitatif et qualitatif des faits sociaux. C'est ce que note Jean-Louis Fabiani quand il souligne le caractère intégrateur de cette approche qui tient à la fois de la démarche analytique et de la démarche narrative [22]. Les méthodes d'appariement optimal constituent un important pour l'analyse de données séquentielles et méritent d'être intégrées à la boîte à outil du statisticien et du sociologue pour être utilisées quand les besoins théoriques l'exigent.

Bibliographie

- [1] Brzinsky-Fay, C., Kohler, U., Luniak, M., « Sequence Analysis with Stata ». *Stata Journal*, n° 6, p. 435-460, 2006.
- [2] Halpin, B., « Sequence analysis utilities for Stata », <http://teaching.sociology.ul.ie/seqanal/>, consulté le 2 septembre 2009.
- [3] Lesnard, L., « SEQCOMP, une extension Stata pour la comparaison de séquences », http://laurent.lesnard.free.fr/article.php3?id_article=7, consulté le 2 septembre 2009.
- [4] Gabadinho, A. et al. (2008) *Mining sequence data in R with the TraMineR package: A user's guide*. Université de Genève, 2008.
- [5] Chenu A., Herpin N., « Une pause dans la marche vers la civilisation des loisirs ? », *Économie et statistique*, n° 352-353, p. 15-37, 2002.
- [6] de Saint Pol, Th., « Le dîner des français : un synchronisme alimentaire qui se maintient », *Économie et Statistique*, n°400, pp.45-69, 2007.
- [7] Abbott A., Hrycak A., « Measuring resemblance in sequence analysis: an optimal matching analysis of musicians careers », *American Journal of Sociology*, Vol. 96, n°1, pp. 144-185, 1990.
- [8] Larmet G., « La sociabilité alimentaire s'accroît », *Économie et Statistique*, n° 352-353, pp. 191-211, 2002.
- [9] Godard F., « Les temps du quotidien », in O. Donnat and P. Tolila (dir.), *Le(s) public(s) de la culture*, Paris, Presses de Sciences Po, 2003.
- [10] Robinson J., Godbey G., *Time for life: the surprising ways Americans use their time*, Oxford, University Park, The Pennsylvania State University Press, 2^e édition, 1999.
- [11] Schor J., *The Overworked American: the unexpected decline of leisure*, New York, Basic Books, 1991.
- [12] Gershuny J., *Changing times: work and leisure in Postindustrial Society*, Oxford, Oxford University Press, 2000.
- [13] Chenu A., « Les horaires et l'organisation du temps de travail », *Économie et Statistique*, n° 352-353, pp.151-167, 2002.
- [14] Durkheim E., *Les formes élémentaires de la vie religieuse*, Paris, Alcan, 1912, Rééd. Presses Universitaires de France, « Quadrige », Paris, 1965.
- [15] Lesnard L., « Schedules as sequences: a new method to analyze the use of time based on collective rhythm with an application to the work arrangements of French dual-earner couples », *Electronic International Journal or Time-Use Research*, Vol. 1, No. 1, pp. 60-84, 2004.
- [16] Lesnard L., « Optimal Matching and Social Sciences », *Document de travail du CREST*, n° 2006-01, 2006.
- [17] Lesnard L., « Flexibilité des horaires de travail et inégalités sociales », dans Insee (dir.), *Données Sociales - La société française*, Paris, Insee, 2006.
- [18] Lesnard, L., « Flexibilité et concordance des horaires de travail dans le couple », dans Insee (dir.), *Données Sociales - La société française*, Paris, Insee, 2006.
- [19] Bouffartigue P., Pendaries J.-R., « Formes particulières d'emploi et gestion d'une main-d'œuvre féminine peu qualifiée : le cas des caissières d'un hypermarché », *Sociologie du travail*, Vol. 36, n°3, pp. 337-359, 1994.
- [20] Prunier-Poulmaire S., « Flexibilité assistée par ordinateur. Les caissières d'hypermarché », *Actes de la recherche en sciences sociales*, n° 134, pp. 29-65, 2000.
- [21] Halpin B., Chan T. W., « Class careers as sequences: an optimal matching analysis of work-life histories », *European Sociological Review*, Vol. 14, n° 2, pp. 111-130, 1998.
- [22] Fabiani J.-L., « Pour en finir avec la réalité unilinéaire. Le parcours méthodologique de Andrew Abbott », *Annales HSS*, n° 3, pp. 549-565, mai juin, 2003.