
**TRAJECTOIRES GÉOGRAPHIQUES ET RÉSIDENTIELLES DES FRANCILIENS
DEPUIS 1950 : ANALYSES DE SÉQUENCE ET ARBRES D'INDUCTION À
PARTIR D'UNE ENQUÊTE BIOGRAPHIQUE**

Guillaume LE ROUX, Arnaud BRINGÉ*, Matthias STUDER**, Christophe IMBERT***,
Catherine BONVALET**

() Institut National d'Etudes Démographiques*

*(**) Université de Genève, pôle de recherche LIVES*

*(***) Université de Rouen, UMR IDEES*

guillaume.le-roux@ined.fr

bringe@ined.fr

matthias.studer@unige.ch

christophe.imbert@ined.fr

bonvalet@ined.fr

Mots-clés : données longitudinales, analyse de séquence, arbre d'induction, enquête biographique

Résumé

Cette communication cherche à mettre en œuvre des méthodes d'analyse de séquences sur des données longitudinales multidimensionnelles et à évaluer leurs apports dans le champ des sciences humaines et sociales, et plus particulièrement, dans l'étude des trajectoires géographiques et résidentielles. En France, en démographie, la mise en place à partir des années 1980 d'enquêtes biographiques, collectant des trajectoires résidentielles, familiales et professionnelles, permet d'étudier les comportements individuels en relation avec les parcours de vie et les interactions entre les différentes dimensions de ces parcours. Les outils de collecte quantitative des parcours de vie se sont progressivement développés (fiche Ageven notamment), et les méthodes d'analyse des biographies se sont nourries du transfert de méthodes depuis d'autres disciplines (informatique, biologie, sciences physiques) vers les sciences sociales. D'abord centrée sur les événements des parcours de vie avec des approches causales [1], les méthodes d'analyse des biographies se développent désormais aussi à partir d'approches « holistes », qui considèrent les parcours dans leur ensemble, et exploratoires : les analyses de séquence [2]. Elles se justifient par l'idée que les trajectoires individuelles doivent être considérées dans leur ensemble, que les événements biographiques ne doivent pas être étudiés séparément mais dans leur succession, et que la relative complexité et la réversibilité de certaines transitions biographiques [3] ne peuvent être appréhendées correctement à travers une approche centrée sur l'événement. Les méthodes développées cherchent alors à décrire la diversité et la complexité des trajectoires individuelles dans leur ensemble et à mettre en évidence des trajectoires-types [4-5]. Une difficulté apparaît lorsque cette approche cherche à intégrer le caractère multidimensionnel des parcours de vie et à en analyser les déterminants. En réponse, différentes méthodes ont été développées telles que les méthodes d'appariement optimal multicanal [6], à visée principalement typologique. S'appuyant sur ces méthodes de calculs de distance entre séquences multidimensionnelles, les analyses par arbre d'induction [7] permettent de mettre en évidence les facteurs qui contribuent à différencier les parcours de vie. L'algorithme de cette méthode cherche à partitionner successivement la population selon les valeurs d'un prédicteur choisi pour différencier le plus possible les partitions produites à chaque étape. Cette communication propose de mettre en œuvre cette méthode sur les données de l'enquête Biographies et Entourage, réalisée en 2001 à l'INED sur la génération 1930-1950 résidant

dans la région Ile-de-France (près de 3 000 parcours de vie) [8]. Nous nous intéresserons aux trajectoires géographiques au sein de l'Ile-de-France - considérées comme des séquences de localisations au sein de la région -, et résidentielles - définies comme des séquences de statut d'occupation (propriété, location, etc) -, entre 20 et 50 ans. L'application de la méthode s'attachera à mettre en évidence les caractéristiques sociodémographiques et les éléments des histoires individuelles (par exemple, l'histoire migratoire et la mobilité sociale familiales) qui discriminent au mieux ces trajectoires conjointes. L'évaluation de la segmentation s'intéressera non seulement au pouvoir discriminant de ces variables et à la stabilité de l'arbre d'induction selon les variables introduites, mais elle s'attachera également à évaluer les logiques « sociologiques » mises en évidence. Pour cela, en s'inspirant de travaux développant des approches mixtes alliant données qualitatives et quantitatives [9], nous mobiliserons un corpus de 140 entretiens approfondis réalisés sur une partie de l'échantillon pour étudier dans quelle mesure la hiérarchie des facteurs mise en évidence par l'analyse quantitative se retrouve dans les discours des enquêtés.

Bibliographie

- [1] Courgeau, D., et E. Lelièvre (1989), *Analyse démographique des biographies*. Paris : Éditions de l'INED.
- [2] Abbott, A. (1990). A primer on sequence methods. *Organization Science* 1(4), 375–392.
- [3] GRAB (2006). *Etats flous et trajectoires complexes. Observation, modélisation, interprétation*. Paris : Éditions de l'INED.
- [4] Studer, M. (2013). *WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R*. LIVES working paper, 23.
- [5] Robette N. (2011). *Explorer et décrire les parcours de vie : les typologies de trajectoires*, Paris : CEPED, collection « Les Clefs pour... ».
- [6] Pollock, G. (2007). Holistic trajectories: A study of combined employment, housing and family careers by using multiple-sequence analysis. *Journal of the Royal Statistical Society A* 170 (1), 167–183.
- [7] Studer, M., G. Ritschard, A. Gabadinho, et N. S. Müller (2011). Discrepancy Analysis of State Sequences. *Sociological Methods & Research* 40(3), 471-510.
- [8] Bonvalet, C., & E. Lelièvre, éd. (2012). *De la famille à l'entourage: l'enquête Biographies et entourage*. Grandes enquêtes. Paris: Ined éditions.
- [9] Duvoisin, A. (2017). *Les origines du baby-boom en Suisse: une approche biographique des cohortes féminines (1910-1941)*. Thèse de doctorat. Université de Genève.