

## CHAPITRE 10

---

# LE TEMPS ET L'ANALYSE DES BIOGRAPHIES

Philippe Antoine

Directeur de recherche à l'ORSTOM, chercheur au CEPED

et

Philippe Bocquier

Chargé de recherche à l'ORSTOM, chercheur au CERPOD

Le temps est l'une des préoccupations majeures des démographes, et le diagramme de Lexis, où le temps est explicité dans ses trois dimensions (âge, période et cohorte), en est la parfaite illustration<sup>1</sup>. Cette perception du temps se retrouve dans l'analyse des biographies, où l'on suit les divers états d'un même individu. Les événements interviennent en fonction de son âge, mais aussi de la période où il risque de connaître un événement. Si, par exemple, on termine ses études à 18 ans, dans une période de récession économique, les chances de trouver un travail sont moindres. Prenant leur origine dans le traitement de données épidémiologiques sur de petits échantillons, les techniques d'analyse quantitative des biographies connaissent maintenant une large diffusion dans toutes les sciences sociales. On parle alors d'analyse des durées de séjours, ou encore d'analyse des événements ou des transitions (*event history analysis*). Plus précisément, lorsqu'on fait intervenir des événements de différente nature dans l'analyse, on parle alors d'analyse des biographies. Les techniques d'analyse des biographies sont de plus en plus accessibles aux chercheurs, même non-statisticiens, mais elles demandent un travail préalable de conceptualisation très rigoureux. Le traitement du temps est

---

<sup>1</sup> Voir à ce sujet Vanderschrick (1992) qui retrace une histoire du diagramme de Lexis, et en cherche le véritable auteur.

absolument central dans ces analyses. Les atouts essentiels de l'approche probabiliste sont la maîtrise des variables explicatives et l'évaluation des interférences entre événements de nature différente. Après avoir présenté les différentes modalités de collecte des événements dans le temps, nous présenterons les avantages et les limites des enquêtes biographiques.

## 1. Reconnaître au recueil rétrospectif ses mérites

Pour mieux saisir les événements démographiques dans le temps, les démographes ont procédé à de nombreuses innovations méthodologiques, en particulier dans les pays où l'état civil est déficient. La plupart des études démographiques traitent d'événements majeurs de la vie de l'individu : mariage, naissance des enfants, décès. Une grande part des acquis méthodologiques a concerné l'observation suivie, soit par la mise en place d'observatoires de population, soit en développant des méthodologies spécifiques comme les enquêtes renouvelées ou les enquêtes à passages répétés (Lohlé-Tart et Clairin, 1988 ; Tabutin, 1984). Dans le cas des observatoires il s'agit de suivre pendant plusieurs années un échantillon de population et d'enregistrer au fur et à mesure tous les événements démographiques. La collecte des données est fiable, mais il faut attendre plusieurs années avant de disposer de résultats montrant l'évolution des comportements. Les enquêtes à passages répétés ont pour avantage de dater plus précisément les événements, car elles font moins appel à la mémoire des enquêtés, en relevant les nouveaux événements entre deux passages espacés en général de 6 mois. Toutefois ces enquêtes ne couvrant souvent qu'une période relativement courte (dépassant rarement une année), ne peuvent fournir que des indicateurs ponctuels, sans profondeur historique, et cela, pour un coût relativement élevé (à peu près proportionnel au nombre de passages).

Les enquêtes rétrospectives ont fait l'objet de multiples critiques, liées notamment aux déficiences de mémoire des enquêtés. Par ailleurs de nombreuses enquêtes rétrospectives recueillent l'ensemble des événements démographiques vécus par les individus enquêtés, mais ne considèrent bien souvent, pour les caractéristiques socio-économiques de ces derniers, que les modalités observées au moment de l'enquête. C'est par exemple le cas des enquêtes démographiques et de santé (EDS). Les enquêtes biographiques au contraire mettent en relation les événements démographiques vécus par l'individu avec son itinéraire professionnel, son itinéraire résidentiel. Le recueil des biographies s'appuie sur un bon repérage dans le temps des événements vécus par l'enquêté. Peu de personnes mémorisent les dates des événements qu'elles ont vécus, mais en revanche, l'enchaînement des événements familiaux est facilement gardé en mémoire. Dans le cas de l'enquête conduite à Dakar (Antoine *et al.*, 1992), pour aider les personnes enquêtées à placer dans le temps les principaux moments de leur vie, nous avons eu recours à la fiche

AGEVEN (âge-événement). Avant de commencer à remplir le questionnaire, l'enquêteur demande à l'enquêté de situer dans le temps les principaux événements de sa vie familiale, puis de sa vie migratoire et résidentielle, et de sa vie professionnelle. Ces événements sont replacés au fur et à mesure de l'entretien sur une fiche où figurent une échelle de temps (années calendaires) et la durée écoulée depuis l'événement. Le classement des événements s'opère ainsi sans trop de difficultés.

## 2. Du recueil des données à l'exploitation statistique

Recueillir des biographies nécessite un questionnaire relativement long et complexe. Les événements sont classés dans le temps, et le plus pratique est de collecter l'information par grand thème : activités, migrations, etc. À chaque changement de profession par exemple l'enquêteur remplit une nouvelle "ligne" d'informations. À l'issue du terrain on se retrouve avec une série d'informations biographiques concernant un même individu. Le plus fastidieux et le plus délicat est de passer du questionnaire aux fichiers informatiques. Dans le cas de l'enquête de Dakar, par exemple, pour chaque individu, nous disposons à la fois de son itinéraire migratoire, de son itinéraire professionnel à Dakar, de son itinéraire résidentiel en ville, de son itinéraire matrimonial, de sa descendance. Pour chaque thème nous constituons un fichier où pour un même individu les événements sont datés et classés dans le temps. Ensuite il faut fusionner ces différents fichiers, afin d'obtenir dans un fichier unique l'ensemble des événements vécus par l'individu. Ce travail informatique exige la définition d'un certain nombre de procédures que l'enquête de Dakar a permis de tester (Bocquier, 1995)<sup>2</sup>.

Chaque enquêté se voit attribuer dans le fichier autant de lignes qu'il a connu de changements d'état. On peut ainsi savoir, à tout instant de sa vie, sa profession, son lieu de résidence, sa situation matrimoniale, ainsi que les renseignements afférents à ces divers statuts, et mettre ces éléments en relation. Ces différentes informations sont datées, et l'on connaît pour chaque période de la vie de l'individu, le temps passé dans un état.

Pendant longtemps les difficultés de programmation et l'insuffisance des moyens de calcul faisaient obstacle à la pratique de l'analyse des biographies. Le développement rapide des capacités de calcul sur micro-ordinateurs et l'amélioration considérable des logiciels statistiques ont rendu les programmes de calculs accessibles. On ne compte plus les logiciels qui, accompagnant les progrès

---

<sup>2</sup> Ces procédures mises au point par Philippe Bocquier ont été proposées aux concepteurs du logiciel STATA qui les ont acceptées et elles seront prochainement diffusées par STATA à ses utilisateurs (STATA, 1994).

de la micro-informatique, proposent des procédures de plus en plus simples pour l'analyse des biographies.

### **3. Le temps, variable déterminante pour l'explication causale**

Pour qu'il y ait causalité, il faut qu'une cause (appelée conventionnellement X) produise un effet Y, sans réciprocité. Si X cause Y, alors Y ne peut simultanément causer X (principe de l'asymétrie causale). Dans cette formulation, le terme "simultanément" est essentiel : c'est parce que le processus se déroule dans le temps que la relation causale est asymétrique. Le temps est à la base de la perception que l'on a de la causalité. Pour percevoir une relation, il est nécessaire de donner à la cause le temps de produire son effet. C'est dès la conceptualisation que l'on doit se poser la question du temps.

Le principe de priorité temporelle de la cause sur l'effet n'est guère contesté. Ce principe épistémologique est accepté par l'ensemble de la communauté scientifique, quelle que soit la discipline. Pourtant, bien des analyses passent outre, parfois même sans que leurs auteurs en aient conscience, faute d'une définition correcte du temps. L'indétermination du temps d'observation est ainsi à l'origine de bien des erreurs d'interprétation. Dès lors que l'on veut mettre en rapport un événement avec un autre, avec telle ou telle caractéristique de la population, on devrait toujours commencer par se donner les moyens de les situer correctement dans le temps, les uns par rapport aux autres, et s'efforcer ensuite de concevoir des modèles d'analyse qui tiennent compte du temps. Chaque fois qu'une caractéristique variant dans le temps est analysée à travers un modèle qui, lui, ne tient pas compte du temps, les résultats risquent fort de conduire à une interprétation erronée de la réalité.

### **4. Population soumise au risque, troncature, risques concurrents : les bases de l'analyse des biographies**

La démarche fondamentale consiste à mesurer le temps qui s'écoule entre un instant de référence commun à tous les individus analysés et la date de l'événement observé ou celle de la sortie de l'observation. Cela nécessite un effort important de conceptualisation. Il faut définir précisément l'événement étudié, la population soumise au risque de le subir, ainsi que les événements concurrents pouvant faire sortir l'individu de l'observation. Si, par exemple, on étudie le passage du premier mariage au divorce chez les hommes à Dakar, la population soumise au risque sera composée des hommes en première union résidant à Dakar depuis le début de leur

union ; le temps qui s'écoule sera mesuré depuis la date de cette union jusqu'à la date de divorce. Toutefois l'observation peut-être tronquée si l'individu quitte Dakar (il émigre avant son éventuel divorce), si son épouse décède (il devient veuf). S'il reste présent et toujours marié (l'individu est donc toujours soumis au risque), la date de troncature sera la date de fin d'observation, c'est-à-dire la date de l'enquête. Ce type d'analyse permet de dépasser l'analyse transversale et de prendre en considération les différents états qu'a connu un individu. Ainsi peut-on par exemple confronter l'itinéraire matrimonial à l'itinéraire professionnel et mesurer les influences de l'un sur l'autre. Le mariage des femmes accélère-t-il ou non leur entrée sur le marché du travail ? Le divorce change-t-il ce rythme d'entrée ? Autant de questions auxquelles il devient possible de répondre.

On calcule pour chaque type d'événement, une série de probabilité par âge de connaître l'événement, que l'on peut combiner pour aboutir à une courbe d'individus subsistants, à chaque durée de séjour, dans un état donné, comparable à la courbe des "survivants" d'une table de mortalité. Cette courbe dite "de séjour", ou encore de Kaplan-Meier, est un des outils exploratoires les plus efficaces de l'analyse des biographies. La courbe décrit le comportement hypothétique d'une cohorte qui n'aurait été soumise qu'au risque de subir l'événement étudié.

### **5. L'avantage des modèles probabilistes : tenir compte des variables explicatives**

Comment passe-t-on de la description des risques à l'analyse de leurs causes ? Le premier stade de la statistique est généralement la construction d'un tableau croisant deux variables. Les résultats de ce croisement sont souvent considérés, à tort, comme "neutres", parce qu'ils sont "simples" à produire. Or, dans un tableau à deux dimensions, on cherchera à expliquer la distribution d'une variable par une autre variable, et par cette autre variable seule. Les variables qui n'apparaissent pas dans le tableau sont supposées ne pas avoir d'effet. Un tableau croisé est donc implicitement un "modèle" postulant l'existence d'une variable explicative unique. On peut compléter le tableau en croisant les deux premières variables avec une troisième, puis une quatrième variable... l'analyse descriptive contiendra toujours un modèle d'explication implicite, même si elle tient compte du temps (comme dans les courbes de séjour), et quel que soit le nombre de variables prises en compte.

De plus, il y a trois inconvénients majeurs à augmenter le nombre de dimensions dans une tabulation croisée. D'abord, l'effet des variables explicatives sur la variable expliquée est combiné : on ne peut clairement identifier l'effet propre d'une des variables explicatives, puisque celles-ci jouent toutes en interaction sur la variable expliquée. Ensuite, un tableau à plus de quatre dimensions est souvent incompréhensible pour une personne normalement constituée (y compris un

statisticien). Enfin, en multipliant les dimensions d'un tableau, on réduit le nombre d'observations dans chacune des cases et nombre d'entre elles perdent toute signification.

Que faire dès lors qu'on ne peut se satisfaire ni des biais d'une tabulation univariée (variable par variable), ni de l'absurdité d'un grand tableau croisant toutes les variables disponibles, dont la plupart des cases n'ont aucune signification statistique ? On fait appel à des techniques d'analyse de régression capables de mesurer l'influence de chaque variable explicative sur la variable expliquée tout en éliminant l'influence de toutes les autres, quel que soit leur nombre. C'est le sens de la phrase rituelle "toutes choses égales par ailleurs".

Rappelons-nous que les tableaux croisés ne sont pas moins des "modèles", au sens statistique du terme, que les modèles de régression. Simplement, les schémas de causalité ne sont pas les mêmes. Dans un cas, les variables explicatives jouent en interactions, dans l'autre, elles jouent indépendamment.

## 6. Le modèle semi-paramétrique dit de Cox

David Roxbee Cox eut la bonne idée de combiner deux types d'analyse : régression et tables de survie (Cox, 1972). On peut voir le modèle de Cox comme l'élimination par la régression des effets d'interaction entre variables explicatives dans l'analyse de survie, ou encore comme l'introduction de la dimension temporelle dans la régression. Les avantages d'une technique permettent de combler les lacunes de l'autre (Cox, 1972 ; Courgeau et Lelièvre, 1989).

Pour résoudre le problème de la durée et des facteurs explicatifs, Cox propose une analyse de régression non pas sur la caractéristique actuelle (au moment de l'enquête) de l'individu, mais sur les différentes valeurs prises par cette caractéristique, durant chaque période (année) de son existence, jusqu'au moment de l'enquête. En quelque sorte, chaque année vécue par chaque personne enquêtée constitue une tranche d'observation. La modalité de référence, exigée par le modèle de régression, n'est pas unique pour l'ensemble de l'échantillon, mais multiple, chaque durée d'observation ayant la sienne. C'est la série construite à partir des probabilités évoquée plus haut, sous forme de "courbe de séjour" qui sert de référence (par exemple, s'agissant de l'analyse du premier emploi, ce sera la série des "subsistants", par durée, dans l'état de "non encore actif", qui fera référence). Cette série, appelée aussi "fonction de séjour de base" constitue la composante non paramétrique du modèle.

Ce modèle de régression mesure alors l'effet des variables explicatives sur le risque annuel de connaître l'événement. C'est la composante paramétrique du

modèle, qui s'ajoute à la composante non paramétrique, pour former un modèle dit semi-paramétrique. À chaque variable est associé un coefficient de régression qui mesure l'influence moyenne de cette variable sur le risque annuel.

## 7. Spécificité de l'analyse des biographies : la résolution des interférences

Les variables explicatives les plus communément utilisées caractérisent l'individu dès sa naissance : sexe, appartenance à un groupe ethnique, à une caste, etc. On suppose que l'effet de chacune de ces variables est constant tout au long de la vie de l'individu ; c'est ce qu'on appelle des conditions permanentes. Leur effet est généralement supposé proportionnel à la probabilité annuelle de connaître l'événement étudié et peut relever de modèles simples.

Le modèle semi-paramétrique permet d'introduire comme variables explicatives des événements survenus depuis la naissance de l'intéressé, et qui ont pu influencer sur ses chances de connaître l'événement étudié. Par rapport au modèle simple où n'intervient qu'une condition permanente (par exemple l'appartenance à une génération), l'analyse s'enrichit de variables explicatives instables, c'est-à-dire d'événements qui peuvent modifier le cours de la vie d'un individu : on approche ainsi de beaucoup plus près la relation causale élémentaire, en respectant le principe de la priorité temporelle de la cause sur l'effet. Par exemple, l'influence de l'itinéraire de formation (études, apprentissage, chômage, etc.) sur la probabilité d'obtenir un emploi peut être analysée en tant que rendement du capital humain, c'est-à-dire comme une variable interne (endogène) au processus d'entrée ou de mobilité sur le marché du travail.

Cette possibilité de faire intervenir des variables explicatives datées est surtout intéressante lorsqu'on veut rendre compte de l'effet de variables externes (exogènes) au processus, tels que : le mariage, l'itinéraire d'une autre personne du ménage (le conjoint, les enfants...), la fermeture d'une usine, un changement de législation du travail, etc. Ces événements explicatifs ont pour particularité de n'avoir pas forcément lieu avant l'événement étudié, au contraire des événements internes au processus.

Avec les variables indépendantes évoluant dans le temps, on tente de suivre le processus au fur et à mesure qu'il se déroule, en s'affranchissant à chaque étape de l'effet des changements de situation survenant dans la vie des enquêtés. Là encore, c'est d'une correcte identification du facteur temps que l'analyse des biographies tire son intérêt.

En somme, le modèle semi-paramétrique de Cox, permet de prendre en compte le temps à deux niveaux : par sa composante non paramétrique (qui tient

compte de l'interruption de l'observation à la date d'enquête), et par l'utilisation de variables indépendantes fonction du temps. On se situe ainsi au plus près de l'analyse causale en construisant des systèmes de relation entre variables où le temps est introduit explicitement (principe de priorité temporelle), et en confrontant ces systèmes de causalité aux données biographiques. On peut assurément parler de "modèles dynamiques" car c'est bien une chaîne causale qu'on tente de vérifier.

## 8. Quelques résultats en guise d'illustration

C'est lorsque l'analyse des biographies porte sur l'interférence entre des événements de nature différente qu'elle porte tous ses fruits. On peut ainsi étudier des phénomènes aussi divers que l'effet des changements matrimoniaux sur la carrière professionnelle (en particulier chez les femmes), l'effet de l'arrivée d'un nouvel enfant sur la survie du dernier-né, l'effet de la polygamie sur le divorce, l'effet de l'entrée dans la vie active sur le départ du domicile parental, l'effet d'un changement de législation sur l'accès au logement ou à l'emploi, etc.

Voulons-nous, par exemple, connaître les facteurs de divorce. Supposons que dans le cadre d'une analyse descriptive classique, nous ayons constaté qu'à Dakar, sur 511 hommes ayant été mariés au moins une fois, 25 sont divorcés au moment de l'enquête et que, parmi eux 16 % sont chômeurs. Peut-on en conclure que le chômage est un facteur de divorce ? *A priori* non, car la situation économique au moment de l'enquête n'est pas forcément celle qui prévalait au moment du divorce. Le recueil biographique nous permet au contraire de savoir que 106 premières unions des hommes se sont terminées par un divorce et l'analyse de montrer que deux facteurs majeurs accélèrent le divorce : d'une part, les situations de précarité (perte de logement et surtout chômage, qui multiplie par 3 le risque de divorcer) et, d'autre part, l'entrée en polygamie (l'arrivée d'une seconde épouse multiplie le risque de divorce par 4). Finalement, c'est plutôt l'arrivée d'une seconde épouse qui, à Dakar, constitue le principal facteur de divorce de la première épouse.

Autre exemple concernant cette fois les femmes, on constate que l'itinéraire matrimonial influe fortement sur l'accès au salariat. Le mariage est un frein à l'accès au salariat : une fois mariées, les femmes voient diviser par 4 leurs chances d'accès à ce type d'emploi, par rapport au moment où elles étaient célibataires. Le divorce leur donne en revanche plus de chance que le célibat : après un divorce rompant un premier mariage, les chances pour une femme d'accéder au salariat sont 2 fois plus grandes que lorsqu'elle était célibataire, et c'est même 7 fois s'il s'agit d'un divorce rompant un deuxième mariage. Sans l'analyse des biographies, nous ne pourrions traiter de ces interférences entre événements.



Dernier exemple, on constate à Dakar, comme dans de nombreuses villes africaines, que l'entrée en union est de plus en plus tardive. On peut se demander si ce retard résulte d'un "effet de génération" ou si d'autres facteurs, masqués par les générations, sont à la base de ces changements concernant l'entrée en union ? L'analyse biographique montre que, même si on élimine l'effet de variables comme l'instruction, la différence entre générations persiste, mais qu'en revanche les variables économiques jouent davantage puisque, effet important de la crise économique, la précarité retarde la première union.

## Conclusion

Nous avons beaucoup insisté sur le temps. Et pour cause, le fil du temps est notre fil d'Ariane : il relie les causes aux effets. L'analyse causale, nœud gordien des sciences sociales ne peut être approchée que si le temps est correctement pris en compte. C'est en ce sens que l'analyse biographique lève une part importante du voile. Certes, les théories générales en sciences sociales ne peuvent évidemment se réduire à des modèles, aussi performant soient-ils ; au niveau des théories dites auxiliaires (spécifiques ou partielles), l'analyse biographique est assurément appelée à jouer un rôle de plus en plus important. Elle donne une impulsion supplémentaire au va-et-vient incessant entre la vérification des théories et la recherche de nouvelles explications.

## Références bibliographiques

- ANTOINE Philippe, BOCQUIER Philippe, FALL Abdou Salam et GUISSÉ Youssouf Mbarguane, 1992. – "Étude de l'insertion urbaine des migrants à Dakar : présentation de la méthodologie d'enquête", in : Éva LELIÈVRE et Claire LÉVY-VROËLANT (éds.), *La ville en mouvement : habitat et habitants*, p. 247-257. – Paris, L'Harmattan, 357 p.
- BOCQUIER Philippe, 1995. – *Manuel d'analyse des biographies pour la micro-informatique*. – À paraître.
- COURGEAU Daniel et LELIÈVRE Éva, 1989. – *Analyse démographique des biographies*. – Paris, INED, 268 p.
- COX David Roxbee, 1972. – "Regression models and life tables, with discussion", *Journal of royal statistical society*, B, n° 34, p. 187-220.
- LOHLÉ-TART Louis et CLAIRIN Rémy (éds.), 1988. – *De l'homme au chiffre*. – Paris, CEPED, 329 p. (Les Études du CEPED, n° 1).
- STATA (éd.), 1994. – "Utilities for survival analysis with time-varying regressions", *Stata Technical Bulletin*, n° 22, november, p. 22-29.

TABUTIN Dominique, 1984. – *La collecte des données en démographie*. – Liège, Département de démographie de l'Université Catholique de Louvain, Ordina Editions, 258 p.

VANDERSCHRICK Christophe, 1992. – "Le diagramme de Lexis revisité", *Population*, n° 5, INED, p. 1241-1262.





CENTRE FRANÇAIS SUR LA POPULATION ET LE DÉVELOPPEMENT



CLINS D'ŒIL  
DE DÉMOGRAPHES  
À L'AFRIQUE

et à

Michel François