L'homogamie d'éducation et de statut social dans la France contemporaine

Milan Bouchet-Valat

Mémoire de sociologie, Master 2 mention Sociologie et Statistique
École des Hautes Études en Sciences Sociales
2011

Sous la direction de Louis-André Vallet

L'auteur tient à remercier Louis-André Vallet et Eva Lelièvre pour leur aide et leurs conseils, ainsi que toutes les personnes ayant participé à la conception, à la collecte et au traitement des enquêtes et bases de données qui ont rendu possible la présente étude.

Sommaire

Introduction **5**

Mesurer l'homogamie : les trois approches contemporaines **9**

I. Études transversales : la voie « classique » 12
Les familles de modèles log-linéaires et log-multiplicatifs 12
Le choix des échantillons 17
Démarches et résultats des études 18
Les études sur la France : problèmes méthodologiques 21
II. Comparaisons internationales 22
Différentes méthodes 22
Des résultats opposés à ceux des études nationales ? 23
Une approche profondément différente des études nationales 25
III. Approches longitudinales 26
Intérêt et difficultés des données longitudinales 26
Une mesure critiquable des variations 27
La modélisation longitudinale 29
Le cas de la France 31
Nouvelles pistes 32

L'homogamie d'éducation et de statut social : évolutions récentes **35**

I. Les données 35
Les variables 35
Reconstruction des couples 37
II. L'homogamie brute : une baisse en trompe-l'œil 41
La question des nomenclatures 41
Une référence : le choix au hasard 42
Résultats 48
III. L'homogamie nette : un affaiblissement réel 51
L'indice de diagonalité nette 52
Modèles log-linéaires et log-multiplicatifs 57
IV. Les transformations de l'espace social 68
L'éducation : nouvelles oppositions, fin de l'hypergamie 69
L'unification de l'espace socio-professionnel 74

Analyse biographique de l'homogamie : scolarité et mise en couple **79**

I. Choix de la classification des diplômes 80
II. Traitement des données 85
III. Les rythmes de mise en couple 87
Approche non paramétrique 87
Modèles semi-paramétriques 88
Modèles paramétriques 91
IV. Homogamie et mise en couple 98

Conclusion **105**

Bibliographie **109**

Introduction

La question de l'homogamie a été relativement délaissée par la sociologie française dans la période récente. Alors que ce thème fait l'objet d'un net regain d'intérêt à l'étranger, tant en Europe qu'aux États-Unis, au Japon et en Corée du Sud, on ne recense en tout et pour tout que trois études empiriques portant sur la France publiées au cours des vingt dernières années (Forsé et Chauvel, 1995 ; Goux et Maurin, 2003 ; Vanderschelden, 2006a), chacune portant sur un aspect et une méthode particuliers. Pourtant, depuis l'enquête classique *Le Choix du conjoint* de l'INED, dirigée par Alain Girard (1964), la sociologie française avait trouvé dans l'étude de l'homogamie un terrain fécond, du point de vue de l'étude de la formation du couple, avec le prolongement des travaux de Girard (Bozon et Héran, 1987, 1988), mais aussi sur le plan de la stratification sociale (Desrosières, 1978 ; Audirac, 1982 ; Vallet, 1986). Si, depuis, les travaux sur l'homogamie se sont faits rares, et si la réflexion théorique qui l'accompagne s'est singulièrement amenuisée, l'intérêt de cet objet semble loin d'être épuisé – c'est, du moins, ce que nous espérons démontrer dans l'étude qui suit.

En effet, même si l'effervescence de travaux des années 1980 est largement retombée, peut-être paradoxalement suite à une critique, sévère mais *in fine* stimulante, dans laquelle François de Singly (1987) remettait en question nombre d'hypothèses qui sous-tendaient ces analyses, ce n'est pas le signe que tout aurait été dit sur la question, ou que ce sujet ne pourrait rien apporter à la sociologie contemporaine. À l'intersection des sociologies de la famille, des réseaux d'interconnaissance et de la stratification sociale, l'étude de l'homogamie apparaît aujourd'hui comme une voie d'accès privilégiée pour l'analyse des tendances contemporaines des sociétés modernes, dans un contexte où les questions de stratification et d'inégalités sont l'objet d'une attention renouvelée, de la part des chercheurs comme de la société.

Au-delà de la mesure des évolutions actuelles, l'homogamie est au cœur d'un débat anthropologique portant sur caractérisation de l'espace social : son enjeu est en effet de savoir si celui-ci est gouverné par une préférence pour la similarité, sur le mode « qui se ressemble s'assemble » (Blau, 1967), par des stratégies largement inconscientes de reproduction (Bourdieu, 1972), ou par une concurrence entre individus pour le « meilleur parti » (Becker, 1991), avec des conséquences profondes sur la conception des relations sociales que la sociologie doit retenir.

Dans ces débats, l'étude sociologique de l'homogamie tire une partie de sa force du dévoilement spectaculaire qu'elle réalise à chaque fois qu'elle insiste sur la force des déterminants sociaux du choix du conjoint, dans une société qui présente le choix amoureux comme le domaine par excellence où s'exerce la liberté individuelle contre les contraintes sociales, explicites ou non (Bozon et Héran, 1987). Représentant peut-être en cela un des sommets de l'entreprise sacrilège de la sociologie, elle vise à considérer objectivement le processus et le résultat d'une « institution » située au cœur de l'idéologie des sociétés libérales contemporaines. Ainsi l'analyse bourdieusienne, qui visait à démystifier la culture légitime en

la replaçant au sein d'un marché des biens culturels, trouve-t-elle un parallèle, peut-être encore plus frappant aujourd'hui, dans l'application de la « froide raison » au domaine tant mis en valeur des sentiments amoureux – parallèle théorique réalisé par Bourdieu lui-même (1979, pp. 109, 167), mais curieusement beaucoup plus travaillé à l'échelle de groupes (Bourdieu, 1972 ; De Saint Martin, 1985) qu'au niveau, macrosocial, qui nous préoccupe (Desrosières, 1978).

À l'instar de la mobilité sociale, en effet, l'homogamie offre un indicateur exceptionnel des proximités changeantes entre groupes sociaux, avec comme particularité de mesurer des associations synchrones, et non plus intergénérationnelles, entre individus et entre classes. Plus encore que l'association parents-enfants (ou plus spécifiquement père-fils), le lien entre conjoints constitue un déterminant fondamental des relations sociales (Haller, 1981), indépendamment des transformations majeures qu'il a connues en même temps que l'institution familiale dans son ensemble. Le choix du conjoint, s'il est pour le sociologue un indice des cercles de relations des deux conjoints, influence aussi la transformation ultérieure de ces cercles, et donc les relations concrètes qu'établissent et maintiennent (ou non) des membres de groupes sociaux différents. Plus encore, à travers l'univers familial créé par une union, ce sont les enfants sur lesquels s'exerce le plus fortement l'homogamie de la génération qui les a précédés, tant la concentration, ou au contraire l'hybridation, d'héritages économiques, culturels et relationnels peut jouer sur les parcours des enfants qui sont issus de cet univers (Mare et Schwartz, 2006).

Si donc les évolutions de la société et des questionnements de la sociologie n'ont fait, à nos yeux, que renforcer l'intérêt d'un renouvellement des études de l'homogamie, le développement spectaculaire, depuis les années 1980, des méthodes d'analyse statistique fondées sur les modèles log-linéaires, log-multiplicatifs et apparentés, impose tout autant l'actualisation et le prolongement des études « classiques », qui n'ont été initiés que très partiellement en France. D'autre part, le recours novateur à des analyses biographiques, dites aussi longitudinales, devrait permettre d'introduire de nouvelles interrogations plus détaillées, comme l'effet de l'allongement de la scolarité sur l'homogamie d'éducation (Blossfeld et Timm, 2003b), tout en dépassant certaines difficultés méthodologiques auxquelles se heurtaient les autres approches. Ces différentes démarches nous semblent devoir aller de pair, un véritable approfondissement de l'état présent et de l'évolution passée des modes d'association entre conjoints gagnant nettement, comme c'est généralement le cas, à conjuguer les niveaux et les contextes d'analyse.

Dans ce qui suit, nous mobiliserons donc à la fois des approches transversales (étudiant l'état de la société à une ou plusieurs périodes) et des approches longitudinales (suivant la biographie des individus), afin de décrire la structure, l'évolution et les processus de formation de l'homogamie d'éducation et de statut socio-professionnel. Dans une première partie, nous dresserons un état des lieux des travaux existants, des théories et des méthodes sur lesquelles s'est fondée l'étude sociologique de l'homogamie ; ce faisant, nous ébaucherons des critiques des techniques et des données utilisées, afin de proposer une courte synthèse des connaissances actuelles, puis de signaler des pistes de recherche à même de remédier aux problèmes rencontrés. Nous appliquerons alors, dans une deuxième partie, ces remarques à l'étude transversale de l'homogamie d'éducation et de statut socio-professionnel en France à partir de plusieurs enquêtes Emploi de 1969 à 2009 : nous étudierons à la fois l'évolution du niveau de l'homogamie et la structure de l'espace social qu'elle permet à chaque période d'observer. Enfin, une troisième partie sera l'occasion d'entrer dans le détail de la formation du couple grâce à l'enquête Biographies et entourage de l'INED, portant sur une génération

de franciliens nés entre 1930 et 1950 ; cette enquête nous permettra de modéliser le rythme de formation des premiers couples afin de mesurer l'influence de la scolarisation sur l'homogamie d'éducation. Nous essaierons ainsi d'apporter une contribution à la compréhension non seulement de l'homogamie, mais aussi de ses causes.

Mesurer l'homogamie : les trois approches contemporaines

L'étude des formes et des tendances de l'association entre conjoints au sein des couples fait partie des domaines relativement classiques de la sociologie. Celui-ci s'est développé en France à partir du constat d'une forte tendance à l'homogamie, menant à une analyse des processus de formation du couple, la célèbre enquête réalisée à l'INED par Alain Girard (1964) ayant ouvert la voie à des approfondissements mêlant interprétations qualitatives et mesures quantitatives (Bozon et Héran, 1987, 1988 ; De Singly, 1982, 1987), d'une grande valeur dans l'élaboration d'une théorie sociologique de l'homogamie et de ses causes. Plus récemment, suivant en cela un mouvement initié par les études de la mobilité sociale, l'attention s'est focalisée sur la mesure et l'analyse des évolutions de l'homogamie dans le temps, et en particulier sur l'existence d'une tendance vers plus ou moins de similarité entre conjoints, mobilisant des techniques statistiques plus complexes.

Avec pour toile de fond la massification scolaire et l'affaiblissement des aspects les plus rigides de la formation du couple – perte d'exclusivité de l'institution du mariage, progression des séparations et des unions libres, réduction au moins formelle du contrôle parental –, les études ont tenté depuis les années 1980, en France mais surtout dans le monde anglo-saxon et en Europe du Nord, de comprendre dans quel sens l'homogamie évoluait dans les sociétés occidentales depuis les générations nées dans l'entre-deux-guerres. Le corpus d'études qui s'est ainsi développé mobilise un cadre théorique relativement unifié, qui vise à vérifier quelques hypothèses simples. On pourrait donc s'attendre à ce que la cumulativité des connaissances soit ici plus forte que dans d'autres domaines de la sociologie. Pourtant, les difficultés de synthèse des résultats obtenus par les nombreuses études successives demeurent assez significatives. En effet, et probablement grâce au relatif consensus théorique, les débats méthodologiques ont été nombreux pour améliorer les techniques statistiques, et ont mené à de grands progrès, avec comme contrepartie une forte technicité et la difficulté à comparer des travaux ne retenant pas exactement les mêmes démarches.

Les questions abordées sont en revanche en petit nombre. Il s'agit tout d'abord d'examiner si les sociétés occidentales contemporaines, rendues prétendument plus ouvertes par la diffusion de l'éducation, se caractérisent bien par une plus faible tendance à l'homogamie que cela n'était le cas pour les générations dont la mise en couple a eu lieu dans l'après-guerre. Néanmoins, deux types principaux d'homogamie sont à distinguer : d'une part, celle liée à l'origine sociale – soit principalement à la profession, la classe sociale et l'éducation des parents –, est supposée décroître, pour laisser en partie place à l'homogamie liée à l'éducation

des conjoints, du fait de l'importance de plus en plus grande prise par la qualification dans le statut social et économique. On voit dans ce rapide résumé des problématiques pointer les questions méthodologiques qui ont justifié le recours à des techniques statistiques sophistiquées.

Premièrement, il est bien clair qu'homogamie d'origine sociale et homogamie d'éducation vont le plus souvent de pair, si bien que la seconde n'est souvent, avancerait une sociologie critique, que le maquillage de la première par l'idéologie du mérite individuel. Deuxièmement, les études pionnières de l'homogamie avaient déjà mesuré l'évolution du niveau de l'homogamie ; elles n'avaient cependant pas réussi à séparer la part de cette évolution qui était due aux transformations structurelles de la population, et celle qui tenait à une modification des comportements matrimoniaux des individus, autrement dit, à la plus ou moins forte perméabilité des groupes sociaux. Une chose est, en effet, qu'une société voie les hommes se marier de moins en moins souvent avec des femmes de statut social inférieur au leur ; une autre, que ce changement soit dû à une conception plus égalitaire des rôles sexués, plutôt qu'à une simple « contrainte de disponibilité » sur le marché matrimonial.

Ces deux difficultés expliquent que l'étude de l'homogamie soit devenue un champ privilégié d'utilisation de modèles log-linéaires, log-multiplicatifs, et d'autres méthodes apparentées, après plusieurs autres tentatives d'élaboration de méthodes *ad hoc*. En effet, ces modèles présentent comme propriété, comme les analyses de régression, de permettre l'étude, dite multivariée, de la répartition de plusieurs variables dans une population, de manière à mesurer les effets de chacune d'entre elles une fois les effets dus aux autres variables contrôlés. Ils permettent ainsi de résoudre notre premier problème en indiquant, par exemple, dans quelle mesure homogamie d'origine sociale et homogamie d'éducation se recoupent, comment ces deux types d'homogamie évoluent l'un indépendamment de l'autre, mais aussi comment évolue la relation qui les lie.

D'autre part, et comme conséquence de cette première propriété, ces modèles ont pour intérêt fondamental de permettre d'évaluer les variations de la relation entre variables une fois tenu compte de leurs répartitions respectives dans la population (propriété qui est celle de l'*odds ratio*) : il est ainsi possible d'éliminer l'influence sur les taux d'homogamie de la taille des groupes et de leur évolution dans le temps. Dans le cas de l'étude de l'homogamie (c'est d'ailleurs la même situation lorsque c'est la mobilité sociale qui est étudiée), les variables principales indiquent les caractéristiques des deux conjoints, et l'interaction entre elles représente les associations entre différents groupes de la population. Il est ainsi possible de mesurer les tendances historiques de l'homogamie de manière *nette*, c'est-à-dire une fois déduits les mouvements dus aux transformations de la structure de la population en termes d'éducation ou de profession.

D'autres thèmes sont fréquemment articulés autour de cette forme de noyau dur de la sociologie de l'homogamie, qui ont trait à des problèmes sociologiques plus larges comme les rôles sexués ou l'importance de l'intérêt dans la détermination des conduites sociales. Le premier thème se retrouve dans l'étude de l'homogamie à travers toutes les asymétries homme-femme qui sont observables dans la constitution des couples : mariage dit traditionnel dans lequel l'homme possède un statut (éducatif, professionnel...) supérieur à celui de sa femme (cas d'hypergamie de la femme, ou hypogamie de l'homme), ou encore couples dans lesquels le premier est plus âgé que la seconde. L'interrogation est triple : les transformations des rôles sexués ont-elles mené à des modifications dans l'association entre conjoints ? à l'inverse, peut-on repérer des signes de transformation de ces mêmes rôles en observant les évolutions des caractéristiques des couples formés plus récemment ? est-il possible

d'examiner si l'élévation du niveau d'éducation des femmes a rencontré des résistances, le mariage traditionnel ne diminuant en fréquence que sous la contrainte, ou a-t-elle été accompagnée par un véritable changement de comportement matrimonial ?

Le second thème est traité à travers l'opposition schématique entre, d'un côté, théories sociologiques expliquant le choix d'un conjoint au sein de son propre groupe social par des mécanismes culturels, aboutissant schématiquement à valider l'adage « qui se ressemble s'assemble » (Blau, 1967 ; Bourdieu, 1979), et, de l'autre, les thèses tentant d'appliquer le modèle de l'*homo œconomicus* au marché matrimonial (Becker, 1991), postulant ainsi une tendance à se marier avec le partenaire qui présente le statut le plus élevé parmi les conjoints possibles. Cette opposition est bien entendu largement fallacieuse, puisque les deux types de comportement peuvent facilement se conjuguer, ce qui se retrouve dans le fait que Bourdieu, qui peut être vu comme l'un des grands promoteurs de la première approche, a accordé une place centrale aux stratégies de reproduction par le mariage (1972) ; mais aussi parce que les deux corpus d'hypothèses peuvent aboutir aux mêmes motifs d'association entre conjoints, chaque groupe ayant tendance, dans la logique de l'intérêt, à se protéger des groupes inférieurs en privilégiant l'entre-soi, ce qui est indiscernable d'une préférence culturelle pour ses semblables – du point de vue du moins d'une table d'homogamie. Des approches plus adaptées ont donc dû être mises en place pour vérifier la validité de ces thèses.

Tous ces thèmes ont pour cadre une autre transformation majeure qui touche à l'objet même des études de l'homogamie : le mariage. Si pour les générations les plus anciennes il n'était pas trop problématique de considérer le premier mariage comme le véritable couple résumant la vie matrimoniale des deux conjoints qui le forment, c'est devenu une approximation de plus en plus fragile au cours du temps. Les études doivent désormais tenir compte du fait qu'une union a de fortes probabilités d'être rompue, et que les conjoints formeront aussi probablement un nouveau couple avec des individus dont les caractéristiques peuvent fort bien différer d'une manière sociologiquement significative de celles des précédents. Mais un couple peut aussi fort bien ne pas se marier avant longtemps, voire se séparer avant le mariage, amenant celui-ci à ne refléter que les unions qui ont connu un relatif succès, et qui pour cela, une fois encore, peuvent fort bien différer de la moyenne des unions.

Pour cette raison, il n'est pas rare que les études retiennent comme champ tous les couples ayant duré plus d'une année. Ce choix ne résout pourtant pas tous les problèmes : si le couple est un phénomène potentiellement évanescent, il devient essentiel de le considérer différemment selon, notamment, sa durée, l'existence ou non d'une cohabitation, ou encore la naissance d'enfants, car rien ne garantit que, dans l'esprit des individus, deux couples aient le même statut, ni qu'ils apparaissent dans les données comme similaires. Il n'est donc plus possible de se contenter de dresser une table unique de tous les couples : il faudrait idéalement disposer de nombreuses informations biographiques, et mettre en place des modèles plus élaborés de compréhension des processus de mise en couple ; au minimum, prendre garde à la nature des couples étudiés et aux générations retenues. Cette difficulté a mené certaines études à privilégier des données portant sur le déroulement de la vie des individus ; d'autres ont simplement dû multiplier les précautions, ou souffrent de ce problème dans la solidité de leurs résultats.

Nous tenterons dans ce qui suit de présenter trois grandes familles d'études de l'homogamie : d'abord, les études sur données transversales, les plus classiques, qu'elles portent sur un pays à travers le temps, ou sur un groupe de pays, étudiés à la fois selon leurs

évolutions historiques individuelles, et selon des différences entre niveaux de développement différents à une même date ; enfin, les approches longitudinales, fondées sur des données biographiques plus détaillées, et qui, apparues assez récemment, restent encore peu développées malgré leur intérêt. Des revues de la littérature sur l'homogamie sont déjà disponibles : celle de Kalmijn (1998) traite des deux premiers types d'études, alors que celle, plus récente, de Blossfeld (2009), privilégie l'approche longitudinale et comparative entre pays. Toutes deux développent plus précisément les questions théoriques que nous ébauchons simplement ici. Enfin, dans une perspective principalement méthodologique, nous mettons l'accent sur les techniques plus que sur le détail des résultats, et sur les limites à dépasser, les failles que l'on peut envisager corriger, plus que sur les apports – que le lecteur se réfère aux revues déjà citées pour un résumé des connaissances sur l'homogamie et nous pardonne le nombre des critiques qui sont formulées par la suite.

I Études transversales : la voie « classique »

L'étude sociologique de l'homogamie, tout en étant un champ riche de travaux quantitatifs et d'interprétations théoriques, a longtemps souffert du manque d'outils de mesure réellement adaptés aux problèmes spécifiques à ce sujet, que nous avons développés plus haut : nécessité d'indicateurs indépendants des marges, possibilité de dissocier plusieurs aspects de l'homogamie. En France, les études quantitatives du début des années 1980 recouraient encore aux simples pourcentages pour décrire l'homogamie, se limitant par là à des descriptions qui, bien que passionnantes, étaient forcées de rester relativement élémentaires, le manque d'un outil direct pour démêler les multiples effets en jeu pouvant rapidement noyer les analyses dans un inextricable réseau de considérations (Audirac, 1982). Cet obstacle n'a pas empêché ces travaux de rester rigoureux et de fournir des résultats intéressants : avec des techniques simples, De Singly (1982) a par exemple montré que le mariage n'est généralement pas une bonne affaire pour les femmes en termes de carrière professionnelle, alors qu'il l'est pour leurs maris.

Le champ a pris un nouveau départ avec l'introduction des modèles log-linéaires, qui sont introduits assez lentement en France, d'abord sous la forme de l'algorithme RAS (Vallet, 1986), puis en association avec d'autres techniques qui n'ont pas été reprises, comme l'indice de diagonalité nette (Forsé et Chauvel, 1995), qui vise à donner une mesure de la force générale de l'homogamie. Avec ces modèles, les travaux sur l'homogamie sont en mesure de calculer une *homogamie nette* qui correspond exclusivement à la tendance des individus à choisir un conjoint possédant certaines caractéristiques plutôt qu'un autre, ce qui ouvre de nouveaux horizons de recherche. En particulier, des phénomènes sociaux majeurs comme l'élévation du niveau d'éducation, la baisse de la part des ouvriers dans la population, et surtout la forte hausse du niveau d'éducation et de l'activité professionnelle des femmes, ne sont plus confondus dans leurs effets avec des variations dans les comportements matrimoniaux.

Les familles de modèles log-linéaires et log-multiplicatifs

Si cet outil statistique s'impose, il prend en revanche des formes variées qui ne sont pas pour rien dans la difficulté à synthétiser démarches et résultats : il existe autant de modèles concrets que d'articles publiés, et nous avons recensé pas moins de cinq types de modèles

statiques, qui se combinent avec (au moins) quatre manières de modéliser les évolutions temporelles. Cette diversité n'est pas entièrement illégitime, puisqu'il est de bonne méthode de construire plusieurs modèles à confronter aux données, sans compter que tous les articles ne mobilisent pas les mêmes variables, et que les méthodes statistiques s'améliorent avec le temps. Nous pouvons aussi nous réjouir de la richesse des modèles désormais utilisés, qui justifie en partie les tâtonnements des articles plus anciens. Il nous semble cependant qu'une certaine unification est possible, ne serait-ce qu'en se forçant à tester tous les modèles « classiques » dans chaque article au lieu de n'en retenir que certains, ce qui rend plus problématiques les comparaisons.

Nous proposons ici un aperçu relativement détaillé des différents types de modèles, puisqu'il nous semble essentiel à la compréhension des convergences et des divergences dans les résultats obtenus. On pourra se référer à Hou et Myles (2008, p. 348) pour un aperçu détaillé et récent des modèles utilisés par les publications anglo-saxonnes, et à Vanderschelden (2006a, p. 41) pour ceux n'apparaissant que dans la littérature française, ainsi qu'aux articles cités avec chaque modèle pour ceux qui n'apparaissent dans aucune synthèse.

– **Modèle de quasi-indépendance** (homogamie simple) (Blackwell et Lichter, 2004 ; Kalmijn, 1991 ; Schwartz et Mare, 2005) : on attribue à la diagonale de la table un ou des coefficients mesurant l'« attraction » pour un conjoint appartenant à la même catégorie sociale. Il est possible de ne recourir qu'à un seul coefficient diagonal, mesurant la tendance moyenne à l'homogamie pour toutes les catégories, ou d'en introduire un par catégorie, voire, dans certains cas, un pour une catégorie qui se distingue des autres, et un pour le reste de la diagonale (Kalmijn, 1991).

  L'inconvénient de ce type de modèle est de ne pouvoir prendre en compte, au-delà d'une logique binaire d'appartenance ou de non-appartenance, les proximités relatives des différentes catégories. Aussi ce modèle est-il souvent utilisé pour sa simplicité, à titre de test et de référence, avant de complexifier la démarche, afin de vérifier que les ajouts ultérieurs améliorent véritablement la qualité des modèles (Schwartz et Mare, 2005). Il est aussi souvent utilisé en combinaison avec les modèles ci-dessous, puisqu'il est généralement nécessaire de tenir compte du fait que les groupes présentent une forte tendance à l'homogamie simple, tout en complexifiant la modélisation des autres types de couples, situés hors de la diagonale.

– **Modèle de franchissement** (*crossings model*) : le principe est de mesurer la « distance sociale » qui sépare les différentes catégories étudiées, mesurée par les plus ou moins fortes probabilités d'union mixte entre ces catégories (Blackwell et Lichter, 2004 ; Johnson, 1981 ; Mare, 1991 ; Qian, 1998 ; Schwartz et Mare, 2005). Ce type de modèle requiert que les catégories soient ordonnées sur une échelle, puisque les distances sont additives. Formellement, ces modèles introduisent un paramètre mesurant la difficulté de franchissement de barrières entre catégories adjacentes ; le coefficient associé à la probabilité d'un mariage mixte est égal à la somme des coefficients associés aux catégories qu'il faut « franchir » pour passer de l'une à l'autre.

– **Modèle de distance** (Kalmijn, 1991) : suivant la même inspiration, on peut varier le modèle en ne supposant plus une relation additive à mesure que des barrières sont franchies : c'est désormais le nombre de barrières franchies qui compte, quel que soit leur niveau. On assigne donc un coefficient à tous les couples mixtes qui franchissent *n* barrières.

D'autres variantes de ce type sont possibles, visant à réduire au maximum le nombre de paramètres nécessaires à une bonne description des données : des hypothèses simplificatrices peuvent être testées. Ainsi, on peut supposer que les distances sur l'échelle sociale croissent de manière linéaire à mesure que l'on s'éloigne de sa catégorie d'appartenance, de sorte que la distance sociale entre les titulaires d'un BEP et d'une Licence serait le double de celle séparant le premier d'un titulaire du baccalauréat. Une relation non linéaire peut aussi être testée, dans laquelle la distance entre deux groupes se mesure à un coefficient élevé à une puissance égale au nombre de barrières franchies.

- **Modèle de quasi-symétrie** (Blackwell, 1998 ; Hou et Myles, 2008) : ce modèle offre une grande finesse de description des données, mais au prix d'un grand nombre de paramètres à estimer. Il peut être utilisé pour mesurer les *odds ratios* correspondant à chacun des différents types de couples possibles, la seule restriction qu'il impose par rapport au modèle saturé étant la symétrie homme-femme des paires de catégories (les couples ouvrier-employée et ouvrière-employé partageant par exemple le même coefficient). En rajoutant un paramètre global d'asymétrie (coefficient d'hypergamie, *cf. infra*), il est possible d'obtenir une très grande fidélité aux données, tout en conservant un nombre raisonnable de degrés de liberté. Ce type de modèle peut notamment servir comme référence afin de vérifier si les simplifications introduites dans les autres modèles présentent un intérêt en termes de rapport entre qualité de prédiction et parcimonie.

- **Modèle d'association lignes-colonnes** (Goodman, 1979) : appliqué une seule fois à l'étude de l'homogamie (Hout, 1982), ce modèle décrit très bien, selon notre expérience, l'association entre conjoints selon le diplôme. Le principe de ce modèle est d'attribuer des scores à chacune des catégories utilisées en lignes et en colonnes, ce qui correspond bien à des classifications ordonnées hiérarchiquement, comme l'échelle des diplômes. En outre, il est possible d'introduire dans le modèle plusieurs associations séparées, qui décriront chacune une dimension de l'espace social étudié, selon un principe analogue à celui de l'analyse des correspondances (Gilula, 1984 ; Goodman, 1986). Un modèle d'association lignes-colonnes à deux dimensions est ainsi, dans nos résultats, celui qui s'adapte le mieux à la description de l'homogamie selon la catégorie socio-professionnelle. Enfin, sous le nom d'association lignes-colonnes homogène, le modèle permet de forcer l'attribution des mêmes scores aux lignes et aux colonnes, supposant donc une symétrie homme-femme qui semble généralement mieux correspondre aux sociétés développées contemporaines (notamment française et américaine).

- **Coefficient d'hypergamie** : tous les modèles déjà cités se voient très souvent rajouter un paramètre distinguant l'un des triangles de la table (inférieur ou supérieur, ce qui revient par élimination au même lorsque la diagonale est libre), autorisant ainsi une asymétrie hommes-femmes. Si l'on suppose que les catégories sont ordonnées, ce paramètre s'applique en effet de manière uniforme à tous les couples dans lesquels le statut de l'homme est supérieur (ou inférieur, selon le triangle choisi) à celui de la femme : il mesure la tendance à l'hypergamie des femmes (dite aussi hypogamie des hommes).

- **Modèles articulant origine sociale et diplôme** (Forsé et Chauvel, 1995 ; Kalmijn, 1991) : quelques articles étudient à la fois l'homogamie d'origine sociale et celle d'éducation. Dans ce cas, c'est généralement la profession des pères des conjoints qui est croisée avec le niveau d'éducation de ceux-ci. Il ne s'agit cependant pas d'un type

de modèle à part, mais plutôt d'une adaptation des types déjà cités, avec comme principale différence l'interaction entre les deux variables de statut. Kalmijn (1991) utilise un modèle de distance (déjà présenté) double, c'est-à-dire autorisant des paramètres de distance à la fois pour l'origine sociale et l'éducation. Forsé et Chauvel (1995) se contentent d'un modèle laissant varier les marges avec le temps, et comparent selon une méthode originale les modèles selon que l'on introduit ou non une interaction entre origine sociale, diplôme et période considérée ; on est donc ici dans le cadre d'un modèle plus fin encore que celui de quasi-symétrie (pas d'obligation de symétrie homme-femme).

Les modèles que nous venons de présenter permettent de décrire de manière statique l'homogamie ; mais afin de mesurer des variations entre périodes, les articles recourent encore à plusieurs solutions différentes. La majorité des auteurs se contentent d'analyser les variations des coefficients ou des *odds ratios* estimés par les modèles, que ce soit au niveau global (probabilité de mariage hétérogame), ou pour différentes paires de catégories, pour différentes distances sociales entre conjoints... Les articles plus récents y ajoutent souvent une modélisation supplémentaire : au lieu d'appliquer un modèle autorisant des valeurs différentes des coefficients d'association pour chacune des périodes comparées, on introduit des hypothèses concernant l'évolution de l'homogamie entre les périodes ; si l'une des hypothèses décrit suffisamment bien les données, on peut alors analyser les valeurs des paramètres obtenus. Ces modèles varient amplement en complexité, du simple test d'une hypothèse à l'estimation de plusieurs paramètres.

- **Comparaison à partir des prédictions** : la méthode la plus simple pour évaluer le sens et l'ampleur de l'évolution de l'homogamie entre deux périodes est de comparer les tables observées à une période avec une table contenant les valeurs prédites par un modèle à partir des données d'une autre période, généralement la précédente ou un point de référence (Vallet, 1986). En utilisant comme étalon des valeurs prédites à partir, non pas d'une autre période, mais d'hypothèses que l'on sait incomplètes, il est possible de mesurer l'importance de certains effets sur l'évolution de l'homogamie : en particulier, on peut attribuer une partie de l'évolution à une transformation des marges (Vallet, 1986 ; Vanderschelden, 2006a), ou à celle du lien entre origine sociale et diplôme (Forsé et Chauvel, 1995). Pour mesurer la force de l'homogamie dans les tables, le pourcentage d'homogamie stricte peut alors être utilisé (Vanderschelden, 2006a), mais aussi d'autres mesures, comme l'indice de diagonalité nette (Forsé et Chauvel, 1995) .

- **Modèles de constance de l'association, de l'hétérogamie et de l'homogamie** (Vallet, 1986 ; Vanderschelden, 2006a) : il s'agit simplement dans la première version de tester l'hypothèse selon laquelle l'association entre les catégories des conjoints n'a pas varié entre plusieurs périodes. On pose un modèle simple tenant compte des marges de la table à chaque période, mais dans lequel l'interaction entre catégories respectives des deux conjoints prend la même valeur pour toutes les périodes.

  Les modèles de constance de l'hétérogamie partent de ce premier modèle fondamental, mais excluent la diagonale de la table (soit en l'ignorant, soit en l'ajustant parfaitement grâce à un nombre de paramètres égal à sa longueur). Une série supplémentaire de paramètres autorise les associations constatées sur la diagonale à varier, mais pas celles sur les autres cases (correspondant à l'hétérogamie) de la table. Le modèle de constance de l'homogamie fonctionne de manière similaire, mais en autorisant le

paramètre à varier seulement en dehors de la diagonale. Dans tous les cas, on vérifie la qualité du modèle pour déterminer si l'hypothèse d'une absence de variation est plausible.

– **Modèle log-multiplicatif à effet de couche** (*layer effect* ou UNIDIFF) (Erikson et Goldthorpe, 1992 ; Xie, 1992 ; Vanderschelden, 2006a)[1] : si les modèles précédents ont été rejetés, on peut mesurer les variations de la force de l'association en multipliant les coefficients associés à chaque paire de catégories des conjoints par un nouveau paramètre, unique, prenant une valeur différente pour chaque période (modèle log-multiplicatif). Ainsi, on suppose que la force de l'association varie de manière uniforme pour tous les types de couples, autrement dit, que la structure des associations reste constante, seule l'ampleur de l'association globale évoluant.

  Il est par ailleurs possible, à la suite de Vanderschelden, de combiner ce modèle avec les précédents afin de mesurer spécifiquement les variations respectives de l'homogamie et de l'hétérogamie ; voire d'appliquer le principe d'une variation uniforme à une paramétrisation quelconque des associations statiques présentées plus haut, et notamment un modèle de quasi-symétrie.

– **Modèle log-multiplicatif à forme de régression** (Goodman et Hout, 1998 ; Vanderschelden, 2006a) : on introduit deux séries de coefficients mesurant l'association entre un couple (ordonné) de catégories : les premiers coefficients, « isolés », mesurent la constance de l'association dans le temps ; les seconds sont en revanche multipliés par un coefficient spécifique à la période, qui varie conventionnellement entre 0 et 1. Cette seconde série de coefficients associés aux couples de catégories mesure ainsi la part de l'association qui varie entre la période où elle est la plus forte et celle où elle est la plus faible ; les coefficients associés aux périodes décrivent, eux, la position de la période en question dans ce mouvement temporel, permettant de mesurer le rythme de l'évolution globale.

  Ce modèle autorise une certaine finesse, en fournissant à la fois une mesure générale des variations, et une mesure de leur ampleur par couple de catégories. En revanche, en l'état, il oblige à introduire autant de paramètres que de cases dans la table d'homogamie, alors que des modèles plus parcimonieux existent. Néanmoins, il est théoriquement possible de le combiner avec des modèles statiques plus parcimonieux présentés plus haut.

  Une des difficultés du modèle tient à l'absence de mesure globale de l'homogamie : les coefficients de la seconde série fournissent une indication sur les types de couples ayant connu le plus de variations dans la force de leur association, mais nous ne disposons d'aucun moyen d'en déduire une évolution globale. Il est d'ailleurs possible que des tendances contraires soient observées pour différents types de couples, auquel cas l'analyse à l'aide de ce modèle devra s'en tenir à un grand niveau de finesse (ce qui représente évidemment aussi un gain dans l'interprétation).

1 Vanderschelden le désigne sous le nom de modèle de variation uniforme de l'association, ce qui est une description plus évocatrice, mais qui présente l'inconvénient de porter à confusion avec le modèle proposé par Yamaguchi (1987), dénommé « *uniform layer effect model* ».

Le choix des échantillons

Même si des différences importantes existent, comme nous venons de le voir, entre les modèles utilisés, une des sources de divergences majeures entre articles tient aux données utilisées. La plupart des articles, qui portent sur le cas des États-Unis, privilégient l'homogamie d'éducation, et incluent rarement des mesures de l'origine sociale, ce qui exclut d'emblée la comparaison avec les travaux, notamment français, qui en tiennent compte.

Mais c'est surtout la définition des cohortes utilisées pour la comparaison qui pose problème, en raison de l'existence de ce que nous désignerons sous le nom de biais d'attrition. En effet, retenir pour l'analyse le stock des couples à une période donnée (mesure de prévalence) implique d'inclure dans l'échantillon les effets des divorces : or, ceux-ci sont plus fréquents pour les couples déjà anciens, mais aussi pour les couples hétérogames. Des variations dans l'âge moyen au premier mariage, dans les taux de divorce, dans les taux d'hétérogamie, facteurs qui sont tous en interaction avec la date de l'enquête utilisée, entrent donc inextricablement dans la prévalence de l'homogamie. De plus, si, comme c'est le cas dans certains articles, une seule enquête est utilisée, la comparaison portant sur des générations découpées selon leur âge, ce biais est redoublé, puisque les individus les plus âgés ont par construction connu plus de séparations que ceux qui viennent à peine de se marier. Lorsque des enquêtes menées à plusieurs dates sont utilisées, ce biais n'est pas *a priori* identifié, mais des variations dans les taux de divorce, par exemple, peuvent influencer les résultats, même si cette influence peut être considérée comme faisant partie de l'objet étudié.

Pour échapper à ces effets, il peut être appréciable de travailler sur des cohortes définies par l'année de mise en couple (mesure d'incidence), mais ces données ne sont pas toujours disponibles. Il importe en tout cas avant tout de disposer de données concernant les conjoints à une date proche de la mise en couple, plutôt que, comme c'est malheureusement le cas dans certains travaux que nous présentons plus bas, d'informations sur les conjoints à la date de l'enquête (les cohortes les plus anciennes ayant eu plus de probabilité de changer de statut depuis leur mise en couple que les plus récentes) ; voire même d'informations sur un conjoint qui n'est pas forcément celui auquel s'intéresse l'analyse (le premier conjoint), mais, le cas échéant, son dernier successeur à la date de l'enquête. Il n'y a donc aucun sens à vouloir recourir à une mesure d'incidence si les données dont on dispose ne concernent pas la date de mise en couple : le biais d'attrition sera tout aussi fort.

Ce problème étant généralement reconnu, le débat dans la littérature porte sur l'ampleur de ses effets. Les articles qui recourent aux mesures de prévalence font plus ou moins explicitement, et avec plus ou moins de justifications, l'hypothèse que le biais introduit est négligeable (*cf.* par exemple Forsé et Chauvel (1995) pour des justifications empiriques de cette hypothèse). En revanche, d'autres articles (Kalmijn, 1991, p. 500) avancent que ce biais a conduit les études à conclure à une stabilité de l'homogamie d'éducation, alors qu'elle aurait augmenté (les cohortes les plus anciennes étant vues comme plus homogames qu'elles ne l'étaient en réalité lorsqu'elles avaient l'âge des cohortes actuellement les plus jeunes). Mais ce point reste controversé : Kalmijn cite Bumpass et Sweet (1972) à l'appui de son affirmation, cet article mettant en avant une plus forte instabilité des couples en cas de différence d'âge importante, ou de différence de diplôme, mais dans le cas ou celle-ci est extrême. Blackwell et Lichter (2004) concluent, eux, à une relative complexité de la relation entre homogamie et stabilité du couple : les couples en cohabitation depuis plus de trois ans ne présentent pas clairement de différences par rapport à ceux qui cohabitent depuis moins de trois ans, ni même par rapport à ceux qui ne cohabitent pas. Janssen (2001) montre quant à lui que l'hétérogamie ne renforcerait les probabilités de divorce qu'en cas d'hypergamie de l'homme, c'est-à-dire lorsque celui-ci possède un statut inférieur à celui de sa femme. En

France et en Suisse, respectivement, des articles déjà anciens de Commaille (1974) et de Kellerhals et al. (1985) montrent que les taux de divorce sont plus faibles parmi les couples homogames, au contraire de ce que De Singly (1987, p. 188) prétend avec ce qui semble bien être une certaine dose de mauvaise foi.

Une manière de trancher le débat a été expérimentée par Schwartz et Mare (2005), en mobilisant à la fois taux de prévalence et taux d'incidence pour plusieurs points dans le temps, et en utilisant des données issues de séries d'enquêtes successives plutôt qu'une seule enquête rétrospective : les tendances observées sont fortement similaires, même si en quelques points des différences mériteraient examen. Il reste que l'exposition des cohortes âgées sur une période plus longue au risque du divorce joue mécaniquement, et qu'il est toujours préférable, lorsque l'étude porte sur des effets conjoints de nombreuses variables, d'éliminer toutes les interférences prévisibles, en recourant aux mesures d'incidence lorsque les données sont disponibles. Les effets liés à des transformations sociales comme le développement du divorce, le recul de l'âge au mariage, le développement de la cohabitation, gagnent à être étudiés séparément, plutôt que comme bruit de fond inévitable des données.

Démarches et résultats des études

L'interrogation principale des travaux publiés à partir des années 1990 concerne l'évolution de l'homogamie d'éducation, soit la tendance à choisir un conjoint ayant le même diplôme (ou le même nombre d'années d'études, selon les approches retenues) que le sien, ou un diplôme proche. Mare (1991) s'est d'abord penché sur les variations de l'homogamie d'éducation aux États-Unis de 1940 à 1987, et a montré d'une part qu'elle avait augmenté de manière continue jusqu'aux années 1970, et d'autre part que cette augmentation était en partie liée au rapprochement entre l'âge du mariage et l'âge de fin des études qui découlait de l'allongement de la durée de celles-ci. La tendance semblait se stabiliser dans les années 1980.

La même année, Kalmijn (1991) s'est intéressé à l'importance relative de l'origine sociale dans l'homogamie, dans le but de vérifier la validité de certaines théories du changement social, qui avancent que les sociétés développées accordent de moins en moins d'importance au statut hérité des parents (ou assigné, *ascribed*) par rapport aux caractéristiques acquises (*acquired*) par l'individu au cours de sa vie ; soit d'un côté l'origine sociale, mesurée par la profession du père de chaque conjoint, et de l'autre l'éducation de ceux-ci. L'étude présente par ailleurs l'intérêt d'une grande rigueur méthodologique, en étudiant des cohortes de mariages conclus sur une période donnée, plutôt que le stock des mariages en cours durant cette période, qui souffre du biais d'attrition décrit plus haut, et ce à partir de deux enquêtes menées à deux dates différentes. On peut cependant regretter que l'éducation des parents et la profession des conjoints ne soient pas mobilisées, au moins dans le but de vérifier l'importance respective de ces facteurs. Le fait que l'on retienne la profession dans le cas du père, mais l'éducation dans le cas des conjoints, semble en effet révéler une hypothèse, pas vraiment explicitée, selon laquelle la profession, indicateur de la classe sociale, serait plus liée au statut assigné que l'éducation, susceptible de mesurer la réussite individuelle ; pourtant, l'origine sociale se mesure aussi par le diplôme des parents, et le statut acquis, par la profession des conjoints, ce qui n'est pas étudié.

Kalmijn conclut en faveur des théories testées, en notant que l'homogamie d'éducation est plus forte que l'homogamie d'origine sociale (ce qui est vrai même lorsque ces deux dimensions sont contrôlées simultanément, la première étant souvent une médiatrice de la seconde, qui agit en partie indirectement), et que l'homogamie d'éducation s'est renforcée entre 1962 et 1973, alors que celle liée à l'origine sociale a plutôt décru. Plus encore, alors que

l'importante augmentation du niveau d'éducation des femmes sur la période explique une partie seulement de l'accroissement de l'homogamie d'éducation, les transformations de la structure professionnelle de la population expliquent en intégralité l'affaiblissement de l'homogamie d'origine sociale : autrement dit, l'homogamie d'éducation s'amplifie en partie à cause de changements dans les comportements matrimoniaux, alors que les comportements sont stables concernant l'homogamie d'origine sociale. En particulier, les titulaires de diplômes du supérieur (*college degrees*) ne se marient presque plus en dehors de leur groupe ; et les individus ayant suivi un parcours universitaire sans décrocher de diplôme sont aussi peu choisis comme conjoints que ceux ayant quitté le système scolaire à la fin du secondaire, ce qui tendrait à montrer que des logiques de statut jouent, et l'emportent sur la simple influence des lieux de rencontre et d'une culture partagée.

L'auteur montre aussi que les homogamies d'éducation et d'origine sociale n'obéissent pas aux mêmes logiques : concernant la première, les individus ont tendance à choisir un conjoint au statut proche, mais pas nécessairement identique au leur ; au contraire, dans le cas de la seconde, la proximité importe peu, c'est une logique binaire d'appartenance ou non au groupe qui joue. C'est un résultat intéressant qui mériterait d'être repris ; malheureusement, peu nombreuses sont les études qui mobilisent à la fois niveau d'éducation et origine sociale.

C'est donc Kalmijn lui-même qui, dans un autre article (1994), approfondit ce questionnement en opposant deux explications concurrentes de l'homogamie : celle qui présente les affinités culturelles comme déterminantes (hypothèse de la similarité), et celle qui soutient que les individus recherchent avant tout un partenaire de statut économique élevé (hypothèse de la concurrence). Dans une inspiration explicitement bourdieusienne, les professions sont classées selon leur niveau de capital économique d'une part (mesuré par le salaire moyen d'une profession), et culturel de l'autre (mesuré par le niveau d'études moyen). Cette analyse, qui reste largement à l'état d'ébauche, indique que l'homogamie culturelle domine assez fortement l'homogamie économique, mais que cette dernière s'est renforcée entre 1970 et 1980, alors que la première reculait légèrement. En outre, l'hypothèse de l'échange du haut statut culturel de l'un des conjoints avec le haut statut économique de l'autre est rejetée : contre toute attente, et même si les effets restent faibles, il semblerait qu'un haut statut culturel soit plutôt un handicap pour les femmes dans la recherche d'un conjoint de statut économique élevé, et réciproquement pour les hommes. L'originalité et le caractère exploratoire de l'approche, ainsi que l'absence de données concernant les périodes antérieures, ne permettent cependant pas véritablement de confronter ces conclusions à celles des autres travaux.

En France, un article de Forsé et Chauvel (1995) a repris la problématique de l'affaiblissement du statut assigné, et tente de la vérifier en comparant deux générations à partir de l'enquête Emploi de 1989. On peut cependant reprocher à cette étude de supposer explicitement que les divorces qu'a connus la génération la plus âgée n'affectent pas de manière significative la mesure de l'homogamie. En se fondant sur deux études, les auteurs affirment que les effets du divorce sont faibles, et qu'ils n'ont pas de raison d'affecter différemment l'homogamie d'éducation et celle d'origine sociale. Mais contrairement à ce qu'ils affirment, et à la lumière des travaux déjà cités sur le problème du biais d'attrition, il nous semble fort plausible que la modification des taux de divorce qui a eu lieu entre les deux cohortes puisse déformer les résultats. L'étude aboutit cependant, avec une méthode (l'indice de diagonalité nette) qui lui est propre, à des conclusions similaires à celles de Kalmijn (1991) : l'homogamie d'éducation prend le pas sur l'homogamie d'origine sociale ; en revanche, aucune progression de l'homogamie d'éducation n'est constatée. Une des

originalités de l'étude est aussi de examiner si des variations dans la relation entre la PCS du père de l'individu et le diplôme de ce dernier expliquent les variations de l'homogamie : seuls 18 % de celle-ci semblent en découler.

L'article plus récent de Schwartz et Mare (2005) a aussi repris ces questions classiques en utilisant des données sur la société américaine couvrant de longues périodes, et en utilisant des méthodes standard, ce qui permet la confrontation avec plusieurs autres études. Les auteurs se concentrent sur des mesures de prévalence, c'est-à-dire sur le stock des mariages en cours à la date de l'enquête, mais contrôlent aussi leurs résultats par des mesures d'incidence, c'est-à-dire sur le flux des mariages conclus à la date de l'enquête, particularité qui constitue une exception dans la littérature. Dans les deux cas, les données proviennent d'enquêtes successives, ce qui permet de connaître le statut des conjoints à une date relativement proche de la mise en couple, évitant ainsi d'introduire un biais d'attrition lié à la durée écoulée entre celle-ci et l'enquête.

Quelle que soit la population étudiée, les conclusions rejoignent celles des précédents articles sur les questions qu'elles ont en commun. L'homogamie d'éducation semble prendre de plus en plus d'importance dans la société américaine depuis les années 1960, malgré un possible ralentissement depuis les années 1990. Les auteurs contredisent en revanche le premier article de Mare (1991) pour la période 1940-1960, sur laquelle ils observent une diminution de l'homogamie d'éducation. Il apparaît que, dans un premier temps, la fermeture croissante des groupes les plus diplômés était compensée en nombre par un mouvement opposé d'ouverture des autres groupes, en lien probablement avec la massification scolaire ; depuis les années 1970, en revanche, les deux extrémités de l'échelle des diplômes se renferment progressivement. Les explications avancées tiennent à la diminution de l'écart d'âge moyen entre la fin des études et le mariage, source d'homogamie d'éducation renforcée ; à l'accroissement des inégalités économiques, dont les variations sont relativement concomitantes avec celles de l'homogamie, en particulier en ce qui concerne les moins diplômés ; à une transformation du rôle des femmes sur le marché matrimonial faisant suite à l'augmentation radicale de leur taux d'activité.

Très récemment, Hou et Myles (2008) ont montré, en s'appuyant sur les exemples des États-Unis et du Canada, que le niveau d'agrégation choisi pour mesurer le niveau d'éducation affectait non seulement l'ampleur de l'homogamie mesurée à une période donnée, mais aussi son sens de variation, ce qui était moins attendu. Dans le cas des États-Unis entre 1940 et 1960, l'utilisation de catégories très détaillées amène à conclure à une augmentation de l'homogamie, alors que des regroupements plus importants fournissent un résultat inverse. En fait, la transformation de la composition de la population en termes d'éducation fait que le regroupement de certaines catégories a des effets différents au début et à la fin de la période étudiée. La solution proposée est d'utiliser des catégories très détaillées, et de ne s'autoriser à fusionner des catégories que lorsque les effets de la fusion sur le niveau mesuré d'homogamie sont faibles. On peut noter que cela justifie d'autant plus le choix fait par Schwartz et Mare (2005) de présenter des *odds ratios* correspondant à différentes barrières sociales entre niveaux d'éducation, qui permet de porter l'attention sur les points où se joue la distinction entre groupes à un moment donné. Mais il demeure essentiel de vérifier en outre que les catégories utilisées reflètent bien la diversité des situations des individus à toutes les périodes étudiées : en particulier, la tentation est forte de négliger les distinctions, aujourd'hui obsolètes, placées tout en bas de la hiérarchie des diplômes, alors qu'elles sont de première importance pour les périodes anciennes.

Les études sur la France : problèmes méthodologiques

Il est difficile de dire si tous les résultats obtenus pour les États-Unis se retrouvent dans le cas de la France, tant les études quantitatives sont rares : outre les deux articles déjà cités, on ne recense que celle de Goux et Maurin (2003), qui recourent à une méthode différente (voir plus bas), et celle de Vanderschelden (2006a). Ce dernier article, récent, réutilise l'approche en termes de modèles log-linéaires que nous venons de présenter, mais malheureusement avec des modèles différents, ce qui rend difficile toute tentative de situer la France dans ce schéma que dessinent les autres travaux. L'auteur vise avant tout à évaluer les variations de l'homogamie sur la période qui va des années 1930 aux années 1990. Pour cela, l'accent est mis sur les paramètres des modèles évaluant ces modifications : sont successivement testés des modèles postulant une variation de la force de l'association entre PCS des conjoints uniforme pour tous les couples de PCS (modèle de « variation uniforme de l'association ») ; puis une association comportant une part de variation spécifique à chaque couple de PCS (modèle log-multiplicatif à « forme de régression ») ; puis enfin deux modèles de constance, respectivement, de l'homogamie et de l'hétérogamie.

Les précédents travaux n'avaient pas recouru à ce type de modèle, mais mesuraient les variations de l'homogamie par l'étude de l'évolution des valeurs des paramètres, que ce soient des *odds ratios* correspondant aux chances de mariage entre catégories, ou des coefficients de distance sociale (Kalmijn, 1991), qui peuvent être représentés graphiquement (Mare, 1991 ; Schwartz et Mare, 2005), ou encore par l'étude des variations de l'indice de diagonalité nette (Forsé et Chauvel, 1995). La démarche de Vanderschelden, en réutilisant des modèles spécialement élaborés pour mesurer des variations (dans la mobilité sociale notamment), apporte des avancées intéressantes sur le sujet. Mais elle ne permet aucune comparaison avec les autres pays, alors que, d'une part, les études des variations de l'homogamie en France font cruellement défaut, et que, d'autre part, les résultats de l'article vont en partie à l'encontre de ceux des autres travaux. L'auteur observe en effet une légère baisse de l'homogamie nette d'éducation pour les diplômes intermédiaires, une stabilité aux niveaux supérieurs, mais une hausse aux niveaux inférieurs ; l'homogamie socio-professionnelle est relativement stable. Dès lors, il est difficile de faire la part entre ce qui tient réellement à une spécificité du cas français, et ce qui tient aux apports des modèles utilisés, ou au contraire aux limites méthodologiques de l'article.

On peut en effet relever quelques problèmes dans la méthode utilisée. Tout d'abord, les données souffrent toujours du biais d'attrition déjà décrit, lié aux plus fortes chances de séparation des cohortes les plus anciennes : une enquête menée à une seule date est utilisée, l'enquête Étude de l'histoire familiale de l'INSEE (1999). Pour la PCS, seule la dernière union de l'interrogé étant prise en compte, on néglige toutes les unions antérieures qui « n'ont pas tenu » ; pour le niveau d'études, les données manquent pour tous les couples séparés au moment de l'enquête ; ces deux restrictions présentent potentiellement, on l'a vu, le risque de masquer des effets significatifs. Les PCS renseignées sont celles des conjoints à la date de l'enquête, limite qui ne se retrouve pas dans le cas du plus haut diplôme obtenu, tant les chances sont faibles qu'il évolue au cours de la vie des individus. Ainsi peuvent peut-être s'expliquer les divergences avec les résultats de Vallet (1986), qui se fondait sur des enquêtes successives, donc ne souffrait pas du biais d'attrition.

Un second problème découle des modèles utilisés, qui retiennent soit une vision assez stricte de l'homogamie, définie comme la diagonale d'une table, soit une vision très détaillée : les mesures de distance sociale présentées plus haut ne sont jamais utilisées. Ceux-ci auraient peut-être permis de synthétiser en quelques coefficients les multiples évolutions que le modèle log-multiplicatif à forme de régression utilisé oblige à commenter individuellement.

Il nous semble donc nécessaire d'accueillir avec prudence les résultats inédits de cette étude, qui placeraient la France dans une dynamique plus égalitaire que les l'Amérique du Nord. Le fait que Forsé et Chauvel (1995) aient déjà conclu à une stabilité de l'homogamie d'éducation pourrait être l'indice que la France connaît bien une trajectoire spécifique – mais ces auteurs recourent à des méthodes très similaires, qui souffrent des mêmes biais sur les données, et des mêmes limites sur les modèles. Ces réserves ne signifient cependant pas que l'ensemble des résultats soient douteux, mais qu'il faut se méfier des tendances de faible ampleur, qui sont facilement affectées par les biais. Nous nous efforcerons donc de reproduire ces questionnements en éliminant les sources potentielles de biais.

II Comparaisons internationales

Il est clairement possible, ainsi que nous l'avons fait dans ce qui précède, de procéder à des comparaisons entre pays à partir de travaux strictement limités à l'un d'entre eux, puisque les méthodes utilisées sont relativement similaires – il est en fait souvent plus difficile de comparer des travaux portant sur un même pays lorsque les méthodes divergent. C'est ainsi que procède Blossfeld (2009) dans une bonne partie de sa revue de la littérature sur l'homogamie, qu'il effectue *« dans une perspective comparative »*. Mais ce sont les travaux qui se fondent sur des bases de données transnationales qui retiendront ici notre attention.

Ceux-ci sont peu nombreux, bien que les premières approches de ce type remontent au début des années 1990 (Smits, 2003 ; Smits, Ultee et Lammers, 1998 ; Ultee et Luijkx, 1990). Ils se sont tous concentrés sur la mesure et l'explication de l'homogamie d'éducation, pour des raisons principalement pratiques : les indicateurs d'origine sociale ne sont pas disponibles partout, et posent surtout d'immenses problèmes de standardisation, que ce soit entre pays pauvres et pays développés, voire simplement au sein de ce dernier groupe. Les systèmes éducatifs sont beaucoup plus similaires à l'échelle mondiale, et permettent d'envisager des mises en équivalence raisonnables, même sur un échantillon très diversifié de pays. Comme dans toute étude quantitative portant sur des contextes sociaux fortement différents, il nous paraît cependant essentiel de rester prudent, et de garder à l'esprit le sens sociologique des catégories étudiées, afin de ne jamais perdre de vue les limites que leur mise en équivalence à travers un échantillon varié de pays impose à l'interprétation. Cette précaution méthodologique n'est pas une simple remarque de principe : les différents découpages qui ont été mobilisés ont parfois abouti à des résultats contradictoires, ce qui nous force à nous poser la question du sens de ce qui est mesuré.

Différentes méthodes

Bien que plus uniformisée au niveau international que celle des statuts socio-professionnels, la mesure du niveau d'éducation des époux n'est cependant pas exempte de difficultés, notamment lorsque des pays de niveaux de développement très divers sont étudiés. Si la première étude comparative de l'homogamie (Ultee et Luijkx, 1990) s'en tenait aux pays industrialisés, les deux suivantes, en incluant des pays en développement, ont dû effectuer des choix afin de créer des catégories qui aient un sens pour tous les pays considérés ; elles ont pour cela recouru à deux approches différentes. La première (Smits, Ultee et Lammers, 1998) mobilise une définition relative au pays, dite *positionnelle*, du niveau d'éducation, dans

l'hypothèse que l'homogamie d'éducation ne se conçoit qu'en relation avec la société à laquelle appartiennent les individus, et que la fermeture des groupes en fonction de leur diplôme dépend donc avant tout du niveau d'éducation des groupes qui les entourent.

La seconde (Smits, 2003) met en œuvre une méthode d'analyse très similaire, mais fondée sur une autre définition, absolue cette fois, du niveau d'éducation, qui modifie assez profondément le questionnement : il ne s'agit plus d'étudier le comportement d'un groupe relatif, mais d'un groupe défini de manière absolue, par un niveau de diplôme, et dont le statut se transforme du tout au tout d'un pays pauvre à un pays riche. La limite étudiée ici (le fait d'avoir terminé ou non des études secondaires) amène à découper des groupes très fortement minoritaires dans certaines configurations nationales, et majoritaires dans d'autres : ces groupes ont dans le premier cas un statut d'élite, alors qu'ils sont beaucoup plus communs dans le second. Cette définition présente un avantage certain, celui d'être robuste : seuls deux niveaux d'éducation sont retenus, et aucune difficulté de définition n'est à craindre ; elle évacue les objections liées à la définition et à la mise en équivalence des niveaux du diplôme auxquelles s'exposait la précédente.

En revanche, il n'est pas certain que cette approche soit celle qui présente le plus d'intérêt : il est bien évident qu'avec la massification scolaire et l'élargissement du groupe supérieur, le comportement matrimonial de ce groupe a toutes les chances d'évoluer. Une fois le sens et l'ampleur de cette évolution identifiés, l'interprétation n'est pas simple : comment rendre compte des rapports qu'entretiennent deux groupes dont les positions relatives, en termes de statut comme d'effectifs, ont totalement changé ? La première approche (dite positionnelle) présentait l'avantage, en supposant que les équivalences entre pays soient valables, de traiter de situations entre groupes comparables, d'un certain point de vue, d'un pays à l'autre. En outre, on a déjà noté que des variations dans la finesse des classifications utilisées peuvent mener à des conclusions opposées quant au sens des évolutions à l'œuvre : il est pour cela essentiel de garder à l'esprit que l'étude de Smits ne mesure qu'un aspect de l'homogamie d'éducation, et que des évolutions en sens contraire peuvent avoir lieu à d'autres niveaux de l'échelle scolaire.

Notons tout de suite que ces études souffrent toutes potentiellement du biais d'attrition, puisqu'elles se basent sur des cohortes d'âge, et non pas de mariage. Dans le cas de comparaisons internationales, les effets de ce biais sont complexes à appréhender à cause des grandes différences qui peuvent exister dans les taux de divorce entre pays. On peut cependant considérer ce problème comme une réalité sociale qui doit être mesurée : elle fait d'une certaine manière partie de l'objet d'étude, même si l'inconvénient est qu'il est impossible de mesurer son effet propre en le séparant de l'homogamie observée en général dans un pays.

Des résultats opposés à ceux des études nationales ?

À cause de ces différences de définitions et d'échantillons de pays, et malgré la continuité liée à la présence d'un auteur commun à presque toutes ces études – Jeroen Smits –, une comparaison systématique de leurs résultats s'avère aussi délicate que dans le cas, développé plus haut, des études portant sur un seul pays au cours du temps. Dans sa revue, Blossfeld (2009) note que les nombreuses variations méthodologiques rendent difficile toute comparaison des résultats des différentes études. Celles-ci se sont en effet focalisées sur différentes manières de mesurer l'homogamie d'éducation : outre les différentes classifications de diplômes qui ont été utilisées, on peut relever le fait que c'est tantôt l'homogamie globale qui est mesurée, tantôt les chances de mariage mixte entre deux catégories voisines. Il est dommage que les auteurs n'aient pas profité du fait qu'ils

travaillaient sur des séries de données qui incluaient, en les élargissant, celles de leurs prédécesseurs pour conduire une analyse restreinte, c'est-à-dire limitée aux données des études précédentes mais utilisant la nouvelle approche proposée, pour examiner si le changement de cadre d'analyse était, seul, à l'origine des différences de résultats. Quelques évolutions schématiques, que les travaux visaient à tester sous forme d'hypothèses, sont cependant repérables ; mais elles semblent, au premier abord du moins, profondément diverger des tendances identifiées par les études centrées sur un seul pays.

La première étude, conduite par Ultee et Luijkx (1990) sur des données décrivant 23 pays industrialisés, conclut à un affaiblissement de l'homogamie d'éducation à travers le temps, et à l'existence d'une relation en forme de « U renversé » par laquelle l'homogamie d'éducation croît dans un premier temps avec la massification scolaire, avant de décroître à nouveau. Smits, Ultee et Lammers (1998) arrivent sensiblement au même résultat général, et remarquent à partir de leur échantillon élargi de pays que deux hypothèses classiques de l'étude de l'homogamie sont vérifiées à des stades successifs du développement. La première, dite de l'obtention d'un statut (*status attainment*), postule que l'homogamie d'éducation s'accroîtrait à mesure que le diplôme devient le déterminant central du statut des individus dans les sociétés développées ; la seconde, dite de l'ouverture générale (*general openness*, ou de l'amour romantique), avance que l'homogamie devrait au contraire décroître à mesure que le niveau de richesse de la société permet aux individus de s'offrir le « luxe » du mariage d'amour, rendu possible par des contacts plus fréquents entre groupes. Il apparaît ainsi que la première hypothèse est vérifiée lorsque l'analyse se concentre sur les pays sous-développés à intermédiaires, alors que la seconde l'est dans le cas des pays intermédiaires à avancés, conduisant *in fine* à l'apparition d'une courbe en « U inversé » sur l'ensemble de la trajectoire de développement d'un pays.

La perspective légèrement différente de Smits (2003), menée à partir de données sensiblement identiques, conclut à une baisse généralisée de l'homogamie d'éducation, aussi bien au sein de chaque pays, entre générations, qu'entre pays, avec leur modernisation. Dans ce modèle, l'indice du niveau de développement (*modernization*) du pays n'a pas d'effet significatif sur l'homogamie d'éducation, et une nouvelle variable, mesurant la taille du groupe des plus diplômés, se révèle être la variable cachée que recouvrait le niveau de développement, ces deux variables étant fortement corrélées. Il ne s'agit donc pas d'une remise en cause des résultats précédents, mais plutôt d'un approfondissement : les sociétés les plus développées ne présentent pas de caractéristique intrinsèque qui expliquerait leur moins forte homogamie ; c'est principalement parce que la fin des études secondaires ne constitue plus une barrière structurante, à cause de la taille trop importante du groupe qu'elle sélectionne, qu'elle voit l'homogamie se réduire à son niveau. Mais les véritables barrières discriminantes peuvent bien s'être reconstituées ailleurs.

Smits pointe d'ailleurs que la réduction de l'homogamie d'une génération à l'autre au sein d'un même pays dépend principalement de la taille relative du groupe des plus diplômés, puis de la vitesse de modernisation du pays. Le bouleversement des structures sociales découlant d'une modernisation rapide expliquerait ce dernier phénomène. Il est assez naturel, même si les auteurs n'avancent pas cette explication, de supposer que les pays dont le développement est plus rapide connaîtront une baisse plus rapide de leur homogamie, simplement parce qu'ils effectuent le même « trajet » que les autres en moins de temps : cet effet ne reflèterait rien d'autre que la transformation accélérée d'une société. Quant à la taille du groupe des plus diplômés, elle est négativement corrélée à la diminution de l'homogamie, de sorte que les pays les plus développés connaissent une évolution beaucoup plus lente que les autres.

Associée aux différentes réserves déjà avancées, cette précision ne peut que nous conduire à être prudents quant à une supposée réduction de l'homogamie au sein des sociétés occidentales : si les pays en développement peuvent bien faire l'expérience d'une forte atténuation des barrières sociales, il n'est pas du tout certain que ce phénomène se vérifie dans le cas de ceux qui ont déjà connu ce processus. Ni même que la tendance, à d'autres niveaux d'éducation et dans des proportions relativement faibles, se soit déjà inversée, comme le suggèrent les travaux plus fins menés sur des pays en particulier. Si la conclusion de ces travaux comparatifs semble en opposition complète avec celles des études nationales, qui ont conclu à un renforcement de l'homogamie d'éducation dans les sociétés développées, c'est donc peut-être que deux niveaux de lecture différents sont utilisés : la contradiction ne serait alors qu'apparente.

Une approche profondément différente des études nationales

Nous pouvons conclure de l'opposition entre études comparatives et études nationales que les différences de méthode sont probablement à la source de la plupart des divergences ; les différentes sources de données fourniraient donc des conclusions similaires, à approche équivalente. Il n'est en effet pas possible de mettre sur un même plan ces deux types d'études, dont les méthodes divergent déjà au sein de chaque groupe sur des points non négligeables : le caractère apparemment erratique des résultats tiendrait donc en grande partie à des différences dans les niveaux d'observation. Les différents travaux, en se concentrant chacun sur un type précis d'homogamie, ne traitent pas toujours du même objet : il importe donc de comprendre comment s'articulent leurs résultats apparemment contradictoires.

Les études concentrées sur un pays, qui portent toutes sur des sociétés développées, mesurent l'homogamie d'éducation selon des classifications diverses – en nombre d'années d'études ou par catégorie de diplôme –, mais recourent toujours à des découpages assez fins pour refléter des tendances qu'on pourrait qualifier de faibles, au regard des différences qui peuvent être observées entre pays de niveaux de développement différents. Les études comparatives se focalisent, au contraire, sur des oppositions entre de larges groupes sociaux dessinés assez schématiquement, et dont on peut aisément concevoir qu'ils aient vu leurs frontières respectives devenir plus poreuses dans les sociétés les plus développées, quand bien même, *à un niveau de classification inférieur*, et qui n'a de sens que pour des sociétés fortement éduquées, des inégalités plus « faibles », mais pas pour autant plus négligeables, se renforcent. C'est ce que suggère Smits (2003, p. 257) à propos des résultats contradictoires obtenus par les études cherchant à mesurer les évolutions de la mobilité sociale – résultats qui ont été mis en cohérence depuis.

Mais ce renforcement peut aussi s'opérer à tous les niveaux d'éducation, pour peu que la tendance à choisir son conjoint dans le même sous-niveau de diplôme se soit accrue en même temps que le choix d'un conjoint dans un niveau sensiblement éloigné soit devenu *moins improbable*. C'est une hypothèse qui ne semble pas avoir été testée dans la littérature comparative existante. Comme nous l'avons vu plus haut, différents types d'agrégation des niveaux d'éducation peuvent mener à des résultats contradictoires.

En outre, les études « nationales » tentent de mesurer l'évolution de l'homogamie dans le temps au sein d'une même société, alors que les études comparatives jouent sur deux dimensions de l'évolution : d'un côté, des générations d'un même pays sont comparées, selon des modalités proches de celles des études nationales ; de l'autre, des pays de niveaux de développement différents sont comparés. Ce dernier aspect, qui constitue la grande originalité de ces travaux, entretient des relations complexes avec le premier, de sorte que la question de

l'évolution historico-temporelle de l'homogamie mérite une approche nuancée. En effet, il est possible d'imaginer qu'au niveau mondial, l'évolution de l'homogamie puisse se décomposer, à l'image de la distinction entre effet d'âge et effet de génération, entre des variations dues au niveau de développement d'un pays, à sa position dans une trajectoire propre, et d'autres qui sont dues à un contexte historique qui affecte tous les pays, ou une partie d'entre eux, dans le sens d'une plus ou moins grande ouverture sociale. Sans prétendre réinterpréter les résultats des trois études déjà présentées à la lumière de cette distinction, il importe de ne pas négliger ce hiatus entre les deux familles d'études pour ne pas opposer leurs résultats de manière artificielle. Nous devons au contraire tenter d'articuler ces deux types d'approche à leurs niveaux différents ; et même les combiner avec un troisième type d'études, fondées sur des données longitudinales.

III Approches longitudinales

Cette troisième démarche, qui n'a été initiée que plus récemment, vise à avoir un meilleur aperçu des mécanismes en jeu dans l'évolution de l'homogamie, en effectuant en quelque sorte un pas en direction des analyses plus qualitatives des processus de formation du couple. En ne se contentant plus de mesurer un état observable à la date de l'enquête (l'existence d'un couple dont les conjoints partagent certaines caractéristiques), mais en entrant, dans une certaine mesure, dans la trajectoire biographique des individus, donc dans les étapes ayant mené à ce résultat, on peut espérer comprendre, et pas seulement constater, le phénomène de l'homogamie.

Par ailleurs, l'utilisation de données biographiques (ou longitudinales) est une manière de résoudre plusieurs problèmes méthodologiques. Tout d'abord, la définition même du concept d'homogamie gagne en rigueur : dans le cas d'une enquête transversale, celle-ci correspond, du fait de l'absence d'autres données, aux caractéristiques du conjoint au moment de l'enquête ; dans une approche longitudinale, il est possible de choisir plusieurs moments dans la vie du couple qui possèdent un véritable sens sociologique : moment de la rencontre, de la mise en couple, de la séparation éventuelle, ou encore de la naissance, du départ du premier enfant. Ensuite, le biais d'attrition disparaît : il est possible de reconstruire la situation matrimoniale de l'échantillon à une date donnée sans risque de déformation ; une autre approche, qui résout aussi ce problème, est d'étudier les couples indépendamment les uns des autres, et de distinguer ainsi effet d'âge et effet de génération dans la mesure des évolutions de l'homogamie.

Intérêt et difficultés des données longitudinales

Les premiers auteurs à avoir développé une approche longitudinale de l'homogamie sont Blossfeld, Drobnič et Rohwer (1998), dans un article disponible en français, qui traite principalement de la question des trajectoires professionnelles des conjoints et des effets qu'elles exercent l'une sur l'autre. Mais c'est surtout l'ouvrage collectif *Who Marries Whom?* (Blossfeld et Timm, 2003b), qui traite de l'homogamie d'éducation à travers 13 contributions couvrant chacune un pays européen, auxquelles s'ajoutent des travaux couvrant les États-Unis et Israël, qui retiendra notre attention ici.

L'ouvrage s'inspire du fait que l'utilisation de données longitudinales fournit une entrée pour l'étude de la formation des couples, permettant ainsi de vérifier l'importance des systèmes éducatifs dans le choix du conjoint, selon l'hypothèse classique (Blossfeld, 2009, p. 518) qu'ils constituent des marchés matrimoniaux privilégiés, pour des raisons liées à la fois aux opportunités de rencontres, à une certaine homogénéité culturelle et à la recherche d'un partenaire de statut social et économique élevé. En amont de l'éducation de l'individu, l'importance des réseaux liés à l'origine sociale est aussi testée, en mesurant l'influence du niveau d'éducation du père sur les « corrections » de trajectoires individuelles, c'est-à-dire aux cas d'hypogamie (mariage « vers le bas » par rapport à son statut acquis) pour les individus ayant connu une ascension sociale, ou au contraire d'hypergamie pour ceux ayant connu un déclassement. Il s'agit aussi de vérifier des hypothèses couramment avancées concernant l'homogamie, comme le renforcement dans les sociétés développées de l'homogamie d'éducation.

L'appréhension du système scolaire comme marché matrimonial n'est pas complètement nouvelle dans les études quantitatives de l'homogamie. Mare (1991, p. 26) avait déjà montré que les individus qui se marient rapidement après la fin de leurs études choisissent plus souvent un partenaire du même niveau d'éducation. Il ne disposait cependant pas de données longitudinales, et avait pour cela dû estimer l'âge de fin d'études en prenant comme approximation celui auquel l'individu aurait dû décrocher son diplôme s'il avait effectué un parcours sans interruption ni redoublement. La parution d'un ouvrage résolument tourné vers les études longitudinales ne signifie malheureusement pas que ce problème, lié principalement à l'existence d'enquêtes adéquates, est résolu. À cause de l'absence de données longitudinales assez complètes dans de nombreux pays, certains articles de Blossfeld et Timm (Blossfeld et Timm, 2003b) ont dû recourir au subterfuge utilisé par Mare : c'est le cas pour la Flandre (p. 47), l'Espagne (p. 153).

D'autres difficultés plus gênantes viennent attaquer la solidité de la démarche. Ainsi, certaines contributions ont dû reconstruire les caractéristiques des conjoints successifs en les supposant identiques à celles du dernier conjoint, qui sont les seules disponibles : c'est le cas de la France (p. 66, voir plus bas), de l'Espagne (p. 153) et de la Grande-Bretagne (p. 177). Cette opération revient malheureusement à réintroduire le biais d'attrition, et à condamner par avance l'un des intérêts principaux de la démarche : mesurer des variations dans le choix du conjoint en fonction du moment de mise en couple. L'article concernant la Suède a dû exclure les 14 % de l'échantillon pour lesquels le niveau d'éducation du conjoint actuel n'était pas disponible, évitant grâce à cette solution de souffrir de ce biais ; mais cette possibilité n'était pas ouverte à tous les pays, pour lesquels trop d'informations faisaient défaut. On peut reconnaître aux auteurs la rigueur de préciser cette limite de leurs études, même si certains prennent plus au sérieux ce problème que d'autres, qui restent peu explicites à son sujet. Certains articles se montrent trop évasifs quant à la nature exacte des données utilisées, de sorte qu'il n'est pas possible de comprendre si elles sont réellement longitudinales ou si elles ont été extrapolées à partir de données transversales, et dans quelles conditions (c'est le cas pour les États-Unis et le Danemark, notamment).

Une mesure critiquable des variations

N'invalidons cependant pas par avance la méthode et les résultats de cette recherche : les lacunes des données ne condamnent pas les techniques qui ont été établies pour cette étude, et l'ouvrage contient des articles dont les données semblent assez solides pour supporter les conclusions des auteurs. L'ensemble des contributions de l'ouvrage est construit sur le modèle

de l'article pilote de Blossfeld et Timm concernant l'Allemagne de l'Ouest (2003a), qui se fonde sur des données véritablement longitudinales et fait preuve d'une grande rigueur dans la présentation des variables et des techniques utilisées. Des variations sont introduites par chaque article, en particulier concernant la présentation du contexte propre à chaque pays, mais aussi en fonction des données disponibles ; certains auteurs ont en revanche tendance à trop diverger par rapport à l'article de référence, ce qui rend la lecture et la comparaison des chapitres plus laborieuses.

La première technique utilisée pour mesurer les variations de l'homogamie est assez simple : elle consiste à comparer le pourcentage des effectifs de la table qui se trouvent sur la diagonale à ceux qui s'y seraient trouvés dans une situation de mariage aléatoire. Ceci revient en fait à utiliser un modèle log-linéaire d'indépendance des niveaux d'éducation du mari et de la femme (intégrant seulement les effets dus aux marges). Appliquée à une série de cohortes définies par année de naissance, cette méthode permet de vérifier si l'homogamie a été plus forte qu'attendue, et, si oui, dans quelle mesure.

C'est sur ce dernier point que la simplicité de cette méthode rencontre ses limites : faut-il comparer les pourcentages attendus et observés selon une méthode additive (écart) ou multiplicative (rapport) ? À titre d'illustration des divergences qui peuvent apparaître entre ces deux façons de calculer, la contribution de référence, portant sur l'Allemagne de l'Ouest (Blossfeld et Timm, 2003a, p. 24), montre un écart croissant entre pourcentages d'hypogamie des femmes attendus et observés, des cohortes 1934-1938 à 1954-1958 (Tableau 1). En revanche, à cause de l'irrégularité des tendances, et de la faiblesse des pourcentages, le rapport entre pourcentages évolue de manière beaucoup moins claire, passant de 2,20 à 3,22 entre les cohortes retenues comme référence par les auteurs, mais de manière très chaotique, et retombant à un niveau très bas à la dernière cohorte (1959-1963). La conclusion des auteurs selon laquelle les probabilités d'hypogamie pour les femmes n'ont pas augmenté comme elles auraient dû le faire d'après les évolutions structurelles du marché matrimonial peut donc être critiquée : les variations observées sont irrégulières, voire inexistantes, et leur sens général reste pour le moins incertain.

Cohorte	Observée (O)	Attendue (A)	Écart (A-O)	Rapport (A/O)
1919-1923	4,0	7,4	3,4	1,9
1924-1928	4,0	8,8	4,8	2,2
1929-1933	4,8	11,0	6,2	**2,3**
1934-1938	**6,2**	10,8	4,6	1,7
1939-1943	5,0	9,9	4,9	2,0
1944-1948	**7,8**	**13,9**	6,1	1,8
1949-1953	4,2	**14,1**	**9,9**	**3,4**
1954-1958	**5,5**	**17,7**	**12,2**	**3,2**
1959-1963	**8,4**	**16,7**	**8,3**	2,0
Moyenne	5,5	12,3	6,7	2,3

Tableau 1 : Différentes manières de mesurer l'hypogamie des femmes

Source : Panel socio-économique allemand (1984-1994), calculs de Blossfeld et Timm (2003a, p. 24) pour les deux premières colonnes ; nos propres calculs à partir de ces données pour les deux dernières colonnes.

On a mis en gras les valeurs supérieures à la moyenne de la colonne.

Nous sommes dès lors forcé d'accueillir avec prudence les résultats présentés à propos de l'évolution de l'homogamie nette, c'est-à-dire des comportements matrimoniaux, une fois déduites les évolutions structurelles. Les variations apparaissent toujours beaucoup plus faibles et irrégulières lorsque les rapports entre pourcentages sont étudiés. D'après ces rapports, il semble que l'hypogamie des femmes se soit affaiblie par rapport à ce qui serait attendu en situation aléatoire, mais la tendance s'est inversée à partir de l'avant-dernière cohorte, née entre 1954 et 1958, au point de revenir au niveau des premières cohortes – ce que l'écart additif ne montre pas. Or, les deux mesures de l'écart et du rapport sont toutes deux connues pour leur fragilité, qui a conduit à leur préférer l'outil plus sophistiqué de l'*odds ratio*.

Plus largement, en effet, l'utilisation de méthodes « non standard » pose toujours un problème à la fois quant aux possibilités de comparaison avec la littérature existante, et quant à la réintroduction de difficultés que des avancées méthodologiques relativement anciennes avaient permis de résoudre. L'utilisation d'un modèle log-linéaire et la comparaison des *odds ratios*, à l'instar de l'intégralité des articles présentés plus haut, aurait évité d'exposer l'ouvrage à ces faiblesses, et aurait épargné au lecteur la charge de vérifier par lui-même les limites de la méthode[2]. Le problème que nous venons d'illustrer sur un exemple présente une difficulté plus générale, puisque la massification scolaire augmente radicalement les pourcentages attendus et observés, ce qui fait mécaniquement croître les écarts additifs, mais pas forcément les rapports et les *odds ratios*. Malheureusement, les données présentées dans l'ouvrage ne permettent pas, en l'état, de recourir à des mesures correctes de l'homogamie. On peut en outre noter qu'aucun indicateur de significativité statistique n'est utilisé, ce qui laisse ouvert le risque de commenter des effets inexistants – et ce risque n'est pas si théorique étant donnée la forte irrégularité des résultats d'une cohorte à l'autre.

La modélisation longitudinale

Nous devons aux auteurs de reconnaître que le cœur de leur démarche et de leur problématique ne tient pas cependant à cette méthode de mesure : c'est bien plus sur le modèle de durée appliqué au long de la vie des individus que l'analyse se concentre. La perspective longitudinale a amené les auteurs à choisir une modélisation en termes de taux de mariage à un moment donné de la vie des individus, avec comme types possibles l'homogamie, l'hypergamie (*upward marriage*) et l'hypogamie (*downward marriage*), soit deux sous-catégories de l'hétérogamie. Un modèle exponentiel de durée vise à tester l'influence d'une série de variables sur les probabilités de se marier selon l'un de ces trois types dans une année donnée. Sont utilisées des variables prenant des valeurs différentes pour chaque année de la vie de l'individu : âge, niveau d'études, interaction entre ces deux variables, poursuite ou non d'études cette année-là, temps écoulé depuis la fin des études ; mais aussi des variables à valeur fixe : niveau d'études du père, numéro de cohorte, interaction entre ces deux variables, et enfin situation de l'individu par rapport au niveau d'éducation de son père (supérieure, inférieure, égale).

Cette méthode aboutit à plusieurs résultats qui semblent solidement établis pour la grande majorité des pays étudiés. Premièrement, la durée du parcours scolaire influe de manière significativement positive sur la probabilité de se marier de manière homogame. Les seuls pays où cette relation n'est pas vérifiée sont les États-Unis et la Suède (effet non significatif)

2 Certains auteurs ont jugé utile de recourir à des modèles log-linéaires, à titre de vérification présentée en note (Heinz et Jonsson, 2003, n. 9 et 12).

et l'Espagne (effet négatif, difficile à expliquer). Cette exception proviendrait d'une plus forte ouverture du système scolaire, qu'il serait facile de suivre jusqu'à 18 ans ; mais une analyse plus approfondie apparaît aussi nécessaire sur ce point.

Ensuite, dans la quasi-totalité des pays, les individus présentent une tendance significative à « corriger » une trajectoire ascendante ou descendante en se mariant plus souvent que les autres dans leur groupe d'origine ; ceux qui ont conservé le même niveau d'éducation que leur père ont une tendance renforcée à l'homogamie. Cet effet est cependant non significatif dans de nombreux pays pour le cas des hommes ayant connu une ascension sociale, ce qui tendrait à indiquer que les hommes ont moins tendance que les femmes à « se laisser rattraper » par leur origine sociale ; mais une véritable vérification de cette intuition devrait s'appuyer sur la comparaison de la valeur des coefficients pour les hommes et les femmes, pays par pays.

Or, on peut regretter que le modèle de régression choisi par les auteurs rende ce résultat particulièrement difficile à lire : au lieu de mesurer l'effet de la relation entre le niveau d'éducation du père et celui de l'enfant sur le niveau d'éducation du conjoint de l'enfant, le modèle mesure l'effet de cette première relation sur les probabilités de réaliser un des trois types de mariage (hypogame, homogame, hypergame). Ce sont donc trois séries de coefficients qui doivent être comparées, sachant qu'un effet négatif sur l'homogamie peut *a priori* indiquer un effet positif aussi bien sur l'hypogamie que sur l'hypergamie : il est nécessaire de se référer en permanence aux trois coefficients pour chaque article. Par ailleurs, il n'est pas possible d'obtenir une mesure de la force de cet effet sur le niveau social *absolu* du conjoint, ce qui limite la portée de l'analyse finale des résultats.

D'autres reproches peuvent être adressés à la modélisation retenue. Le plus général d'entre eux tient au caractère « continuiste » de l'approche : la logique longitudinale incite fortement à recourir à des mesures linéaires, que ce soit de l'éducation ou de l'âge. Or, il est difficile d'imaginer que ces variables, même enrichies en utilisant des écarts à des âges de référence et des logarithmes, ait un effet uniforme sur toute la vie, fournissant un unique coefficient. On s'interdit de cette manière de distinguer différents groupes d'individus pour lesquels les effets seraient éloignés, voire opposés, sans compter qu'une même durée d'éducation peut correspondre à des statuts très différents selon les filières éducatives suivies (Breen et Jonsson, 2000). Cela est d'autant moins compréhensible que la variable mesurant le temps écoulé depuis la sortie du système scolaire donne lieu dans le modèle à pas moins de sept variables dichotomiques (*dummies*), traitement privilégié dont ne bénéficie aucune autre variable.

L'attribution d'un seul coefficient aux variables mesurant le niveau d'éducation de l'individu et de son père a sans doute d'importantes conséquences sur l'interprétation. Il aurait pour cela été utile de savoir si le choix d'attribuer plus de paramètres à la variable déjà citée résulte de l'essai de plusieurs modèles ou d'une décision *a priori*. En effet, une plus grande finesse de la variable reflétant le niveau d'éducation de l'individu, et, plus encore, de son père, au lieu d'une valeur continue (le nombre d'années), aurait eu l'effet de mettre l'accent sur le statut social, relativement stable, au détriment des déterminants conjoncturels. Ne pas laisser la possibilité à ces deux aspects de se confronter aux données risque de forcer les coefficients, alors que l'on sait que, la corrélation de ces variables étant très forte, les effets peuvent passer d'une variable à l'autre assez facilement selon les modèles. L'ajout ou non d'interactions entre variables n'est jamais justifié, alors qu'on aurait pu penser que d'autres interactions que celles qui ont été utilisées auraient pu faire sens : par exemple, celle entre durée écoulée depuis la sortie du système scolaire et niveau d'étude, qui reflèterait les différences de comportement entre des groupes différemment éduqués.

On peut encore noter que les coefficients des trois régressions identiques menées en parallèle sur les trois types de mariage mesurent seulement l'influence des variables sur le taux du type de mariage concerné. Pourtant, c'est bien souvent la part relative de ce type de mariage dans les mariages conclus à un moment donné qui nous intéresse. Avec la mesure utilisée, une croissance du taux de mariage général (homogame comme hétérogame) un an après la fin des études se traduit par un coefficient positif pour cette modalité de la variable, et pour les trois types de mariage ; mais cette information nous importe peu, si, dans la perspective des auteurs, c'est le rôle du système éducatif sur l'homogamie qui nous intéresse : le coefficient positif nous confirme seulement que les mariages sont plus fréquents après la fin des études, ce qui est pour le moins banal.

Le cas de la France

La contribution à l'ouvrage portant sur la France, celle de Goux et Maurin (2003), constitue à notre connaissance le seul article français contemporain qui mobilise une approche longitudinale de l'homogamie. Mais cette contribution est relativement décevante, puisque les données longitudinales utilisées sont en fait en partie reconstruites par extrapolation de données transversales. Ce procédé, dont on a déjà décrit plus haut les limites, entraîne de sérieux doutes quant à la validité des résultats, tant les hypothèses que les auteurs doivent faire pour reconstruire des données longitudinales sont fortes : partant des données de l'enquête Formation et qualification professionnelle (FQP) de 1993, seuls sont disponibles le plus haut diplôme de l'interrogé et de son dernier conjoint au moment de l'enquête. Il est donc nécessaire de supposer que les remises en couples se font avec des conjoints de niveau d'éducation égal, alors que l'intérêt de l'approche longitudinale était justement de postuler que d'importantes variations pouvaient apparaître à ce niveau en fonction de la période de la vie où l'individu se met en couple : on en revient à supposer nul un effet dont l'étude devrait vérifier l'existence et l'ampleur. Seule la date de première mise en couple est connue directement, information qui permet *in fine* de faire tenir toute cette fragile démarche.

Il faut certes reconnaître un mérite à cette étude, qui est celui d'exister, alors que les enquêtes véritablement longitudinales en France ont longtemps fait défaut. Par contraste, des contributions comme celles portant sur l'Allemagne de l'Ouest ou les Pays-Bas peuvent s'appuyer sur des panels de ménages donnant accès aux informations biographiques d'environ 2000 individus. Il nous semble néanmoins que la reconstitution abusive de données longitudinales pose plus de questions qu'elle n'en résout, tant, comme l'histoire des recherches quantitatives sur l'homogamie l'a montré, des problèmes méthodologiques peuvent mener à des résultats erronés. Ce problème est bien réel, puisqu'au sein même de l'ouvrage de Blossfeld et Timm, certains résultats de l'article sur la France détonent, et apportent de la confusion, étant potentiellement attribuables, à notre sens, aux hypothèses qui ont dû être faites pour reconstruire les données.

Ainsi, la France serait le seul pays dans lequel le taux d'hypogamie des femmes était supérieur à ce qui est attendu en situation de mariage aléatoire pour les cohortes les plus anciennes, mais pas pour celles nées après 1949. Elle serait aussi le seul pays dans lequel on observe un effet négatif de l'ascension sociale des hommes et femmes sur leur hypogamie, autrement dit une absence d'« effet de correction » par lequel les individus ayant atteint un niveau d'éducation supérieur à celui de leur père se marieraient plus souvent malgré tout dans leur groupe d'origine : au contraire, en France, ces individus se marieraient moins souvent que les autres avec un partenaire moins éduqué. Ceci pourrait être lu comme le signe d'une plus grande fluidité de la société française ; mais il faut noter que l'effet correspondant (absence

d'effet de correction) s'observe aussi dans le cas des individus ayant un parcours descendant, qui sont ceux qui ont le plus tendance, dans ce modèle, à se marier avec un conjoint moins éduqué, ce qui renforcerait encore leur déclassement.

Ces deux résultats semblent difficilement acceptables, si l'on considère que la variété des pays étudiés dans l'ouvrage est assez importante pour qu'on puisse s'attendre à les vérifier aussi en France. L'existence d'une hypogamie très forte pour les femmes des cohortes nées avant la Libération n'est pas un résultat général des études françaises de l'homogamie : Vanderschelden (2006a, p. 47) identifie par exemple une évolution strictement inverse, les couples dans lesquels l'homme est le plus diplômé l'emportant sur ceux où c'est le contraire qui est vrai au sein des cohortes anciennes, situation qui s'inverse pour les cohortes récentes. Il exige donc une vérification précise de la situation des cohortes en question, d'autant que l'on ne peut que s'étonner que l'hypergamie des hommes (hommes dont le diplôme est inférieur à celui de leur conjointe) ne soit que de 21,7 % sur les cohortes étudiées, alors que l'hypogamie des femmes (la même situation, vue de l'autre côté), qui devrait largement refléter la première (si l'on excepte les couples dont un des conjoints est issu d'une génération trop ancienne ou trop récente, donc non étudiée), soit de 29 %. Une simple étude des profils des couples homogames pour les cohortes d'avant-guerre serait déjà d'un grand intérêt pour expliquer ce résultat.

Nouvelles pistes

Plus récemment, des études longitudinales mobilisant de nombreux aspects biographiques des individus ont été menées aux Pays-Bas, dans le cadre de la thèse d'Ellen Verbakel (2008). Ces travaux représentent une avancée considérable, puisqu'ils éliminent le biais d'attrition, et qu'ils recourent à des analyses multidimensionnelles très riches associant origine sociale, niveau d'éducation, profession, situation matrimoniale et familiale (enfants), le tout dans un cadre biographique permettant la mesure des évolutions tant au niveau de l'individu et du couple que de la société dans son ensemble. Malheureusement, et pour des raisons peu claires, dans leur communication (2007), Verbakel et De Graaf excluent de leur échantillon les individus qui ne sont pas en couple au moment de l'enquête. Ceci réintroduit, sous une forme atténuée, les problèmes de biais d'attrition, ce que les auteurs reconnaissent. En outre, afin de mesurer l'homogamie socio-professionnelle, les couples dans lesquels un des deux conjoints est inactif à au moins une des périodes considérées sont exclus de l'échantillon pour toutes les périodes suivantes. Cette décision, qui aurait pu être évitée simplement en considérant l'inactivité comme un statut professionnel, entraîne de complexes discussions concernant les effets qui pourraient potentiellement en découler, quand la méthode présentait au départ l'avantage d'une grande systématicité ; elle s'explique néanmoins par l'intérêt particulier que les auteurs accordent à la participation au marché du travail, qui est étudiée plus en détail ailleurs, suivant la gradation inactif/temps partiel/temps complet.

La modélisation retenue consiste à appliquer une série de modèles log-linéaires sur l'échantillon à différentes périodes ; plus précisément, plusieurs sous-échantillons sont étudiés séparément, de manière à ne comparer entre eux que les couples qui ont été retenus (*cf. supra*) pour le même nombre de périodes. Les périodes sont ici définies selon, d'une part, la durée écoulée depuis la mise en couple (origine, cinq ans, dix ans, quinze ans), et, d'autre part, l'histoire familiale (origine, un an avant la naissance du premier enfant, un an après celle-ci, entrée à l'école du dernier enfant). Cette approche originale pèche cependant par un aspect fondamental, celui de la mesure de l'homogamie : l'utilisation de l'indice ISEI (*International Socio-Economic Index of Occupational Status*) pour classer sur une échelle numérique les

statuts professionnels permet certes de résoudre les problèmes de catégorisation, mais elle suppose, de même que Blossfeld et Timm (Blossfeld et Timm, 2003b), la continuité des statuts socio-professionnels, alors que les études transversales ont montré que l'homogamie différait grandement en fonction de la force de barrières sociales qui se déplacent dans le temps.

Cet indicateur permet en revanche de tester simplement des hypothèses, une fois acceptée la validité de l'échelle utilisée. Les modèles appliqués à cette variable continue montrent ainsi que le niveau d'éducation du mari a un effet positif sur la profession de sa femme après 10 ans de vie en couple, et après l'entrée du dernier enfant à l'école (mais pas aux autre périodes). La réciproque n'est pas vraie, même si le statut professionnel de la femme un an avant la naissance du premier enfant a un impact positif (significatif mais faible) sur celui de son mari. Ces résultats illustrent l'intérêt de la méthode, mais ils restent trop irréguliers (les auteurs concluent à des effets de faible ampleur), et sont non significatifs pour la plupart des périodes et des variables (l'échantillon de plus de 2000 personnes est considérablement réduit si l'on se restreint aux couples que l'on peut suivre sur toutes les périodes).

La thèse finale (Verbakel, 2008) cherche en premier lieu à mesurer l'influence réciproque des caractéristiques des conjoints sur leurs parcours professionnels. Elle ne recourt pour cela pas souvent à des analyses qui s'accordent avec le point de vue retenu ici. Notons cependant qu'elle conclut (Chapitre 2) à une stabilité de l'homogamie socio-professionnelle (mesurée sur seulement trois catégories), quand le document de travail (2007) constatait une légère et irrégulière diminution, en se fondant sur l'indice ISEI, qui est continu. Comme le note la thèse, le recours à un fort niveau d'agrégation au lieu d'une mesure numérique affecte fortement les résultats, d'autant que les variations observées sont dans tous les cas faibles. Retenons cependant que ces travaux suggèrent une autre modélisation applicable aux données longitudinales, et avancent une manière de tenir compte de l'influence de certaines étapes marquantes d'une biographie. Ce type d'approche peut être mobilisé, *mutatis mutandis*, pour comprendre les processus qui mènent à la formation de couples homogames, et les variations de ces processus au cours des générations.

Au terme de ce parcours à travers la littérature, nous pensons avoir identifié trois voies à suivre pour améliorer la mesure et la compréhension de l'homogamie dans une perspective française. Concernant les enquêtes transversales, qui sont les plus facilement réalisables en raison de la nature des données issues des grandes enquêtes de la statistique publique, il semble possible et nécessaire de compléter les articles récents en mobilisant d'une part les modélisations plus fines qui n'ont pas été appliquées au cas français, et en tentant d'autre part d'éliminer le biais d'attrition par l'utilisation d'enquêtes successives et par l'étude de cohortes de mises en couple. Cette démarche est potentiellement extensible à d'autres pays développés, notamment en Europe grâce aux *Labour Force Surveys*, réunion au niveau européen des équivalents des enquêtes Emploi françaises.

Dans le domaine des approches longitudinales, il est aussi possible de réaliser d'importants progrès grâce à la publication récente d'enquêtes sur échantillons bien plus réduits, mais autrement plus riches que les données transversales, et qui n'ont jusqu'ici pas été utilisées. D'une part, l'enquête Biographies et entourage de l'INED fournit une entrée riche sur les parcours éducatifs, professionnels et matrimoniaux de la « génération » des franciliens nés entre 1930 et 1950, à partir d'un échantillon représentatif de plus de 2000 personnes et de leur

entourage. Si elle ne permet pas de mesurer les évolutions de l'homogamie sur une très longue période, elle offre en revanche un aperçu exceptionnel sur les processus de (re)mise en couple d'une génération dans son contexte historique.

D'autre part, l'enquête SHARELIFE (*Survey of Health, Ageing and Retirement in Europe* – partie biographique), conduite par l'Université de Tilburg sur un financement de la Commission européenne, contient le même type de données, avec pour avantage de couvrir 12 pays européens, en interrogeant près de 19 000 personnes (2000 pour la France), presque exclusivement âgées de 50 ans et plus en 2008 (soit une génération relativement similaire à celle de Biographies et entourage). Ces deux enquêtes sont devenues disponibles très récemment, et promettent de nouvelles perspectives pour l'étude de l'homogamie, en particulier en France, mais aussi pour des comparaisons internationales.

Dans la suite, nous mobiliserons la première approche, transversale, à partir de la série des enquêtes Emploi, puis la deuxième, longitudinale, grâce à l'enquête Biographies et entourage, en nous cantonnant au cas de la France.

L'homogamie d'éducation et de statut social : évolutions récentes

À la lumière des critiques formulées dans la partie précédente à l'encontre des travaux existants, notre objectif dans cette étude sera de mesurer de la manière la plus solide qui soit l'homogamie et ses évolutions dans la France contemporaine. Le biais d'attrition représentant la principale limite de la grande majorité des études de l'homogamie, il était nécessaire de disposer soit de données rétrospectives assez précises pour donner des informations sur les couples en cours à chaque année étudiée, soit de plusieurs enquêtes transversales décrivant les couples à la date de l'enquête. Les enquêtes rétrospectives françaises, on l'a dit, ne remplissent pas ces conditions : l'enquête Étude de l'histoire familiale de 1999 ne donne d'informations que sur les couples en cours à la date de l'enquête ; l'enquête Biographies et entourage (que nous utilisons dans la troisième partie suivant une approche biographique), si elle fournit bien les informations nécessaires, concerne des générations nées entre 1930 et 1950, et ne permet donc pas d'étudier les évolutions récentes de l'homogamie[3]. Nous avons donc emprunté la seconde voie, en utilisant la série des enquêtes Emploi, qui commence en 1968 (pour les données informatisées). La dernière enquête disponible étant celle de 2009, nous avons retenu une enquête tous les 10 ans depuis 1969, incluant donc 1979, 1989 et 1999.

I Les données

Les variables

Nous nous attacherons à décrire deux dimensions de l'homogamie, les plus étudiées : l'éducation et le statut socio-professionnel. La première est mesurée en recourant à plusieurs classifications afin de se prémunir contre les effets, déjà soulignés plus haut, de regroupements arbitraires ; le second, par la classification française des PCS, et par la classe définie selon le schéma international Casmin, qui offre un autre éclairage, et facilite les

3 Cette limite peut cependant être repoussée en étudiant, d'un côté, les enfants des interrogés, qui avaient à la date de l'enquête l'âge d'être en couple, et, de l'autre, leurs parents, ce qui nous permet potentiellement de remonter au début du siècle, époque jamais étudiée du point de vue de l'homogamie (ce sera l'objet d'un travail ultérieur).

comparaisons avec d'autres pays. Les PCS, créées en 1982, ne sont étudiées qu'à partir de cette date (raison pour laquelle nous avons parfois retenu l'enquête de 1982 en plus d'une enquête tous les 10 ans). En revanche, la classe Casmin a été reconstituée sur toute la série depuis 1969, en se fondant sur la classification des CSP avant 1982, sur celle des PCS après, et dans les deux cas sur le code détaillé de profession, la qualification, le statut et le nombre de salariés (le cas échéant)[4]. Ainsi, des trois mesures retenues, deux (diplôme et PCS) sont directement reprises des enquêtes Emploi (après harmonisation).

En revanche, la classe Casmin, parce qu'elle est reconstituée, est d'un maniement plus délicat : dans ce qui suit, nous avons été forcés de nous restreindre aux seuls actifs (chômeurs inclus), à cause de variations importantes dans les enquêtes anciennes de la définition de la dernière profession occupée par les inactifs. Le taux d'inactivité (féminin, surtout) étant très important en 1969 et en 1979, ce problème ne pouvait être négligé, ce qui a justifié l'exclusion de ces individus litigieux de l'échantillon ; mais à l'inverse, il est clair que la sous-population des actifs ne reflète qu'un aspect de l'homogamie. Ce problème est cependant limité de deux façons. Premièrement, en comparant les résultats obtenus pour les seuls actifs, puis pour toute la population, sur la période récente (sur laquelle la définition des inactifs est cohérente), nous avons pu vérifier que les écarts sont toujours faibles, et n'affichent jamais une variation dans un sens différent. Cette comparaison est aussi réalisée, après 1982, avec les PCS, pour lesquelles nous avons toujours repris, pour les inactifs, la catégorie correspondant à la dernière profession exercée.

Deuxièmement, il ne faut pas négliger le fait que, la moitié des femmes entre 35 et 54 ans étant inactives encore en 1975 (Vallet, 1986, p. 662), le statut socio-professionnel des inactifs prend un sens qu'on peut légitimement considérer comme différent de celui qu'il prend pour les actifs : un poste d'employée occupé quelques années par une jeune femme attendant de se mettre en ménage n'a que peu de valeur pour définir la position sociale d'une femme au foyer trente ans plus tard. Pour les périodes anciennes, l'étude des seuls actifs fournit une indication qui, bien que différente, n'en a pas moins du sens. Idéalement, afin de résoudre cette question, une étude devrait en fait croiser statut socio-professionnel des actifs, de la population totale, et origine sociale des individus, afin de vérifier si les tendances concordent.

La mesure du niveau d'éducation présente beaucoup moins de difficultés. Si le choix d'une nomenclature est toujours sujet à débat – nous y reviendrons –, il n'est pas trop difficile de reconstituer le diplôme des individus selon une nomenclature homogène sur toute la série. La classification la plus détaillée que l'on puisse retenir, plus grand commun dénominateur des enquêtes Emploi depuis 1969, est décrite dans la première colonne du tableau 2 ; elle semble assez raisonnablement détaillée pour que l'on n'éprouve pas le besoin d'aller plus loin dans la précision. C'est cette classification que nous utilisons le plus souvent, puisqu'elle présente l'avantage de ne reposer sur aucune hypothèse de regroupement de certains diplômes ; les deux autres nomenclatures sont utilisées à titre de vérification, et de comparaison avec d'autres travaux. Comme l'ont montré Hou et Myles (2008) (*cf. supra*, page 20), le choix d'une classification est d'une très grande importance, non seulement pour la mesure de taux d'homogamie bruts, mais aussi lors de l'utilisation de modèles log-linéaires et log-multiplicatifs. En recourant à plusieurs nomenclatures, nous nous sommes prémunis contre le risque d'obtenir des résultats valables à un seul niveau d'agrégation.

Par ailleurs, la mobilisation de modèles qui ne prennent pas seulement en compte l'homogamie stricte, mais la distance entre les groupes (ce qui est le cas des modèles à effet de couche, des modèles de franchissement et de distance, et des modèles à association lignes-

4 Ce délicat recodage a pu être réalisé grâce au travail déjà effectué par Louis-André Vallet sur les enquêtes Formation et qualification professionnelle, simplement adapté par nos soins aux enquêtes Emploi.

colonnes), permet de limiter l'influence de la nomenclature sur les résultats, puisque le modèle tiendra naturellement compte du fait que deux groupes très proches pourraient éventuellement être fusionnés. Dans la suite, lorsque les différentes classifications amènent aux mêmes conclusions, un seul résultat est généralement présenté.

Nomenclature détaillée	Nomenclature agrégée	Forsé & Chauvel*	Vanderschelden**
Aucun	Aucun	Aucun	Aucun
CEP[1]	CEP	CEP	École primaire
BEPC[2]	BEPC	Secondaire	Collège, CAP, BEP
CAP/BEP[3]	Pro inférieur au Bac	-	-
CAP/BEP avec BEPC	-	-	-
Bac pro/tech/BP/BT[4]	Bac pro/tech/BP/BT[4]	-	Lycée
Bac général	Bac général	-	-
1er cycle[5]	Supérieur	Supérieur	Études supérieures
2e et 3e cycle[6]	-	-	-
École	-	-	-

Tableau 2 : Les quatre nomenclatures de diplôme utilisées

** Forsé et Chauvel (1995). ** Vanderschelden (2006a).*

[1] *Certificat d'études primaires.*

[2] *Brevet d'études du premier cycle du second degré, Brevet des collèges.*

[3] *Certificat d'aptitude professionnelle, Brevet d'études professionnelles*

[4] *Baccalauréat professionnel ou technique, brevet professionnel ou brevet de technicien.*

[5] *Premier cycle du supérieur (inférieur à Bac+2 inclus).*

[6] *Deuxième et troisième cycles du supérieur (au-delà de Bac+3 inclus).*

Reconstruction des couples

À chaque enquête, nous n'avons retenu qu'une partie des cohortes d'âges représentées : aux âges les plus jeunes, les rythmes de mise en couple étant très différenciés selon les groupes sociaux (niveau d'éducation notamment, *cf.* la troisième partie), les catégories se mettant en couple le plus tard seront sous-représentées parmi les couples ; aux âges les plus avancés, la mortalité différentielle entre groupes sociaux, déterminant notamment le veuvage, posera le même problème, sans parler de la mauvaise qualité du codage de la profession des retraités. Nous avons donc choisi de ne conserver dans notre échantillon que les couples dans lesquels l'homme était âgé d'entre 30 et 59 ans à la date de l'enquête : à 30 ans, les taux de mise en couple ont presque atteint leur sommet[5], et à 59 ans, l'individu n'est que rarement en retraite (en particulier pour les enquêtes anciennes, les plus problématiques du point de vue de la qualité des données). Si les limites d'âge n'ont été imposées que sur l'homme (choix arbitraire) plutôt que sur les deux conjoints, c'est qu'il est important de conserver dans l'échantillon des couples présentant un fort écart d'âge, et qui sont peut-être de caractéristiques différentes des autres. Forcer les deux conjoints à entrer dans l'intervalle

5 En 2009, par exemple, 78,4 % des trentenaires vivaient en couple, soit le plus fort taux de toutes les tranches d'âge ; ceci est valable pour toutes les enquêtes à l'exception de 1982, où les quadragénaires les dépassent très légèrement (Tableau 4, page 40).

d'âge retenu aurait éliminé, en partant de chaque extrémité de l'échelle des âges, une proportion décroissante des couples, proportion qui a des raisons de varier au fil des enquêtes – notamment puisque l'écart moyen d'âge se réduit (Vanderschelden, 2006b).

L'enquête Emploi a comme unité d'interrogation le ménage ordinaire (excluant les logements collectifs, comme les hôpitaux, les foyers de travailleurs, les cités universitaires, les maisons de retraite), des renseignements étant obtenus à propos de tous les individus âgés d'au moins 15 ans, soit par interrogation directe, soit en interrogeant un autre membre du ménage. Cependant, s'il est facile d'identifier les individus appartenant à un même ménage[6], il est plus difficile d'associer deux conjoints. Une première restriction, assez évidente, tient à ce que nous n'obtenons que les couples cohabitants à la date de l'enquête. Cela n'est pas un problème, les couples non cohabitants étant probablement très peu nombreux, en particulier aux âges retenus[7] ; l'avantage de cette définition est aussi d'éviter d'avoir à spécifier des critères de durée du couple, ou de demander aux individus de juger eux-mêmes si leur couple est « sérieux » ou non. C'est cette définition du couple que nous retenons par la suite, sans accorder de statut particulier au mariage.

Cependant, dans les données, les liens entre les membres du ménage ne sont pas décrits : seule l'est la relation de l'individu avec la personne de référence du ménage : dite « chef de ménage » dans les enquêtes anciennes, elle correspond à l'homme en cas de présence d'un couple, au parent en cas de famille monoparentale, ou encore, à défaut, à l'homme actif le plus âgé du ménage. Nous avons donc retenu la définition *a minima* d'un couple par ménage, qui fonctionne relativement bien. Nous disposons en effet d'un moyen de juger de la validité de cette approximation : en 1969 et en 2009, il est possible d'identifier plusieurs couples par ménage, ce qui est impossible dans les autres enquêtes. Les variantes b des enquêtes 1969 et 2009, qui reconstruisent le plus finement possible les couples à ces dates, nous fournissent donc un aperçu de l'approximation ainsi réalisée : elle revient à laisser échapper 3,2 % des couples en 1969, mais seulement 58 en 2009 (moins de 0,01 %), ce qui indique que la cohabitation de plusieurs couples dans un même ménage, déjà faible, a presque disparu sur la période étudiée. L'approximation que nous sommes forcés de réaliser introduit donc un biais relativement faible. En outre, même si ce biais devait affecter plus fortement certaines couches de la population, il semble peu probable que les individus ainsi évincés possèdent, à caractéristiques étudiées fixées, un comportement différent du point de vue du choix du conjoint[8].

6 Encore que la documentation des enquêtes anciennes soit lacunaire au point que l'identification d'un ménage grâce à un identifiant unique, qui exige la concaténation de plusieurs variables techniques, a parfois été laborieuse.

7 La définition du couple utilisée par la série des enquêtes Famille de l'INSEE et de l'INED ne retenait, jusqu'en 2011, que les couples cohabitants.

8 Ces hypothèses pourraient évidemment être étayées par une étude du profil des individus qui sont dans le cas décrit, ainsi que par le recours à des mesures alternatives de l'homogamie sur les échantillons « corrects » de 1969 et 2009 (variantes b), vérifications que nous n'avons pas réalisées.

Enquête et variante	Adultes	Ménages	Conjoints ou Couples	Conjoints/ Adultes (%)	Conjoints/ Ménages (%)
1969a	111 103	54 224	38 162	34,3	70,4
1969b	111 103	55 600	39 390	35,5	70,8
1979	116 812	60 210	40 627	34,8	67,5
1982	120 596	63 240	42 322	35,1	66,9
1989	123 742	65 472	42 617	34,4	65,1
1999	136 159	75 276	45 858	33,7	60,9
2009a[1]	72 866	41 724	23 778	32,9	57,5
2009b[1]	72 866	41 724	23 836	32,9	57,5

Tableau 3 : Récapitulatif des individus, ménages et conjoints retrouvés dans les enquêtes

Champ : adultes de plus de 20 ans ; tous les ménages ordinaires ; conjoints cohabitants de la personne de référence du ménage, sauf pour les variantes 1969b et 2009b, pour lesquelles tous les conjoints cohabitants sont pris en compte.

Note : il n'est pas ici possible de restreindre ce décompte aux seuls couples dans lesquels l'homme est âgé de 30 à 59 ans, comme nous le faisons dans la suite. En effet, on ne peut vérifier que l'on n'a pas omis une conjointe tant que son conjoint n'a pas été retrouvé, puisque c'est l'âge de ce dernier qui déterminera l'inclusion ou non du couple dans l'échantillon.

[1] L'échantillon retenu ne correspond qu'à une partie des effectifs de l'enquête Emploi 2009 : par erreur, nous n'avons retenu que les individus interrogés au premier trimestre. Cette restriction n'introduit aucun biais dans les résultats, mais diminue sans nécessité la précision des estimateurs : elle sera corrigée dans des travaux ultérieurs.

On a pu vérifier que tous les individus identifiés comme conjoints de la personne de référence du ménage dans chaque enquête donnent bien lieu, après traitement, à la reconstruction d'un couple (dans le tableau 3, le nombre de conjoints est donc strictement égal au nombre de couples). En 1999 et en 2009, cependant, nous avons exclu de l'échantillon un petit nombre de couples homosexuels, qui étaient totalement absents jusque-là, probablement parce que les enquêteurs ne laissaient pas ouverte la possibilité de déclarer en couple deux personnes de même sexe. Même si cela correspond à de faibles effectifs (170 couples en 1999, 220 en 2009), nous avons préféré les sortir de l'analyse, pour des raisons encore une fois de cohérence avec les enquêtes précédentes, et parce qu'il n'est pas impossible que ces couples présentent un comportement différent en ce qui concerne le choix du conjoint. Les étudier à part n'est malheureusement pas possible à partir des enquêtes Emploi, étant donné leurs effectifs.

En plus de servir de vérification, le tableau 3 nous permet d'observer une légère baisse, depuis 1982, du taux de personnes en couple parmi les adultes de plus de 20 ans. Mais c'est surtout le nombre de couples (ou de conjoints) rapporté au nombre de ménages qui diminue, perdant près de 13 points depuis 1969. Le tableau 4 montre un ralentissement de la mise en couple, perceptible chez les 20-29 ans, et à l'inverse une augmentation du taux de personnes en couple chez les plus de 60 ans. Entre ces deux tranches d'âge, le célibat a eu tendance à se développer légèrement.

Enquête	Âge à l'enquête					
	20-29	30-39	40-49	50-59	60-69	≥ 70
1969	65,5	84,9	84,4	75,2	60,3	30,8
1979	66,7	83,7	83,0	79,0	63,8	35,2
1982	66,8	82,9	85,2	78,5	66,5	36,0
1989	58,8	84,3	81,5	79,5	69,2	36,1
1999	51,9	80,9	79,7	77,8	72,2	44,2
2009	47,9	78,4	77,3	74,8	71,4	49,9

Tableau 4 : Taux de personnes en couple par âge à chaque enquête (en %)

Champ : tous les couples hétérosexuels cohabitants.

Une fois les couples reconstitués pour chaque enquête, il est assez immédiat de calculer des tables d'homogamie croisant caractéristiques de l'homme et de la femme, et de mesurer l'homogamie brute, c'est-à-dire celle qui est observée à chaque période sans autre traitement statistique. Avant d'analyser ces évolutions, le tableau 5 présente les cohortes décennales qui entrent dans le champ des échantillons de couples retenus pour chacune des enquêtes. On peut noter qu'à chaque période (excepté le cas de 1982, utilisé comme point de départ pour les études concernant les PCS) une cohorte de 10 ans en remplace une autre. Les variations observées entre deux enquêtes sont donc le résultat de la différence de comportement entre ces deux cohortes, à laquelle s'ajoutent les remises en couple intervenues dans l'intervalle au sein des autres cohortes. Le premier effet l'emporte cependant sur le second, puisque les remises en couples entre 30 et 60 ans restent *a priori* peu nombreuses sur une période de 10 ans, en comparaison du remplacement d'environ un quart de l'échantillon.

Enquête	Année de naissance							
	1910-19	1920-29	1930-39	1940-49	1950-59	1960-69	1970-79	Toutes
1969	6 280	9 248	8 828					24 356
1979		8 123	8 261	9 379				25 763
1982		*5 566*	*8 146*	*9 572*	*3 347*			*26 631*
1989			7 654	9 122	10 355			27 131
1999				8 545	10 424	9 602		28 571
2009					4 973	4 959	3 941	13 873

Tableau 5 : Effectifs des échantillons de couples par enquête selon l'année de naissance de l'homme

Champ : couples hétérosexuels cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

II L'homogamie brute : une baisse en trompe-l'œil

Mesurer l'évolution de l'homogamie brute relève largement de la gageure. Avancer que dans environ 30 % des couples en 2009, les deux conjoints étaient de même niveau d'études, alors que cette part était de 45 % en 1969 n'aurait en effet pas grand sens si cette affirmation ne s'accompagnait pas de multiples précautions qui privent, on va le voir, cette variation *a priori* assez notable de (presque) toute sa valeur. Si l'on présente ici, de manière paradoxale, des chiffres condamnés à être rapidement disqualifiés, c'est qu'ils permettent néanmoins de situer les mesures *nettes* de l'homogamie, plus robustes, que nous analyserons dans un second temps, dans une réalité dont elles extraient les tendances les plus sociologiquement significatives. Il n'est pas question de nier une certaine valeur aux taux bruts d'homogamie : en 2009, 7,7 % des couples étaient bien formés d'un homme et d'une femme non diplômés ; seulement 2,4 % des couples étaient bien, la même année, formés de deux individus dont le plus haut diplôme était un baccalauréat. Ces chiffres ont évidemment un sens et reflètent des situations sociales concrètes qui ont une réalité et des effets sur la société ; ils font partie de l'histoire des groupes considérés. Ils ne disent pourtant rien, ou si peu, sur la structuration de l'espace social en termes d'ouverture et de fermeture, de hiérarchies, de distinction.

La question des nomenclatures

La première question que soulève l'utilisation de taux bruts d'homogamie tient tout simplement à la classification des diplômes qui est utilisée pour calculer ces taux. Mécaniquement, une classification très détaillée produit des taux plus faibles, et vice-versa : par exemple, l'utilisation de la classification agrégée décrite plus haut plutôt que la classification détaillée fait passer le taux global d'homogamie d'éducation en 2009 de 30 à 37 %. Plus grave : ainsi que nous l'avons déjà souligné (*cf.* p. 20), le sens de l'évolution de l'homogamie peut aussi changer en fonction des catégories retenues. En effet, si une classification n'est jamais neutre, et si elle est d'une importance décisive dans la mesure de l'homogamie, elle constitue l'alpha et l'oméga des taux bruts d'homogamie. Ceux-ci ne sont absolument pas robustes au regroupement de deux catégories : créer des catégories fines au bas de l'échelle donnera une plus grande importance aux variations qui affectent cette partie de la population ; à l'inverse, regrouper ces catégories rendra invisibles certains changements, et donnera plus de poids à des évolutions intervenues ailleurs dans l'échelle.

Ce phénomène n'est pas vraiment un problème si les catégories utilisées ne font pas discussion, ce qui est notamment le cas lorsqu'elles possèdent une très forte réalité sociale, c'est-à-dire si elles sont utilisées par les acteurs sociaux eux-mêmes. Par exemple, concernant les mariages entre Blancs et Noirs aux États-Unis dans les années 1950, les deux catégories utilisées sont clairement définies, toute personne ayant une ascendance noire étant considérée comme noire. Il en va de même dans le cas des mariages entre personnes de religions différentes, l'appartenance religieuse étant (généralement) définie et sanctionnée institutionnellement.

Ce schéma s'applique-t-il à notre étude ? On sait que la nomenclature des catégories socio-professionnelles fait l'objet d'importants débats quant à son adéquation à la réalité sociale. Même si elle tient compte des classements qu'effectuent eux-mêmes les acteurs sociaux[9], et même si on lui reconnaît un grand intérêt pour la description de la société, il est difficile de

9 Sur ces questions, voir notamment l'article original de Desrosières, Goy et Thévenot (1983), ou, pour plus de détails, l'ouvrage plus développé de Desrosières et Thévenot (2002).

présenter la classification des PCS comme un schéma figé. Et l'effet performatif que possèdent les catégories sociales institutionnalisées (Bourdieu, 1984, p. 7) s'accompagne par définition d'un arbitraire des délimitations qui, parce qu'il ne peut jamais épouser parfaitement la réalité des conditions objectives, n'en est jamais qu'une approximation à un moment donné : des catégories peuvent éventuellement éclater en plusieurs, des groupes se détacher de leur catégorie, à l'image des ouvriers et employés non qualifiés (Chardon et Amossé, 2006), et c'est justement ce que doit mettre en évidence l'analyse. Ainsi, le taux d'homogamie brut de chaque PCS ne donne d'informations que si l'on considère qu'il a une réalité propre ; si l'on tient en revanche compte du fait que les limites des catégories ont toujours une part d'arbitraire, ou du moins de discutable, le taux d'homogamie brut perd presque tout son intérêt, puisqu'il n'est absolument pas robuste à la modification, même légère, de celles-ci.

Cette question est encore plus problématique lorsque l'on a affaire à des classifications organisées hiérarchiquement, ce qui est le cas du diplôme. Il ne semble pas que l'ensemble des individus partageant le même diplôme constitue un véritable groupe social, définition impliquant une certaine homogénéité, une conscience de soi et une délimitation nette des frontières du groupe. Les classifications des diplômes couramment utilisées reflètent principalement le fonctionnement du système éducatif, qui sanctionne l'obtention de titres reconnus par l'État, et ignore les trajectoires déviantes, notamment dans le bas de l'échelle : un individu sortant du système scolaire sans avoir décroché un brevet des collèges sera aujourd'hui considéré comme sans aucun diplôme, même s'il est allé jusqu'à la Seconde, au même titre qu'un autre qui, quelques décennies auparavant, n'aurait jamais décroché son certificat d'études, tout en étant inévitablement passé par l'école primaire. Les effets institutionnels liés à la possession d'un titre n'annulent pas ceux, culturels, qui découlent de fréquentations communes et d'un univers culturel proche.

La proportion importante de sans diplôme observée au sein de la population (17,2 % encore en 2009, soit la plus importante catégorie de diplôme dans notre classification détaillée) augmente mécaniquement le taux d'homogamie brut de cette catégorie, et de la population en général. La nette baisse des effectifs de cette catégorie (-30 % en quarante ans) entraîne donc aussi une baisse de l'homogamie brute de la population. Il est cependant difficile de savoir si cette diminution doit être considérée comme réelle (des individus se mettent plus souvent en couple à travers des barrières de diplôme), ou comme un artefact découlant du fait que de plus en plus d'individus rentrent dans les cases définies comme légitimes par le système scolaire et la classification contingente que nous utilisons. Étant donné le luxe de détail qu'offre la classification des diplômes au-delà du baccalauréat, l'élévation générale du niveau d'éducation a pour effet immédiat d'abaisser le taux d'homogamie.

Une référence : le choix au hasard

En effet, le second problème des taux bruts d'homogamie tient à leur fragilité par rapport à la composition de la population en termes des catégories utilisées – même si l'on accepte l'hypothèse selon laquelle celles-ci sont valables. Bien qu'il soit tentant de prendre pour argent comptant la valeur en pourcentage du taux d'homogamie, et d'interpréter une valeur de 30 % comme correspondant au tiers inférieur sur une échelle de l'homogamie qui varierait de 0 à 100 %, cette interprétation n'a pas de sens : une telle échelle n'existe pas. Il est d'une part impossible d'observer une situation dans laquelle tous les couples seraient homogames : à partir du moment où les populations des hommes et des femmes sont différentes du point de vue de la caractéristique étudiée, certains individus devront nécessairement prendre un

conjoint hors de leur catégorie. Cela donne lieu à une homogamie maximale, qui est par exemple de 85,6 % (complément à 100 d'une hétérogamie minimale de 14,4 %) en ce qui concerne le diplôme en 2009 (ce que nous verrons tableau 10, p. 46). Dans l'autre sens, tous les individus ne pourraient pas, même s'ils le voulaient, trouver un conjoint appartenant à une catégorie différente de la leur : on observe donc une homogamie minimale, qui est le plus souvent nulle, sauf lorsque le nombre de catégories utilisé est peu important, et le déséquilibre entre hommes et femmes fort.[10]

C'est pourquoi on recourt à une fiction, qui sert de situation de référence dans l'étude de l'homogamie : la situation de choix aléatoire du conjoint. Au lieu de se fonder sur la position du taux d'homogamie entre le minimum et le maximum, qui pose de nombreux problèmes de dépendance aux marges de la table[11], on compare l'homogamie réellement observée à celle qui aurait eu lieu si les caractéristiques étudiées n'avaient eu aucune influence sur le choix du conjoint. Ce calcul, qui correspond simplement, pour chaque type de couple, au produit des proportions correspondant aux catégories d'appartenance de l'homme et de la femme dans leurs populations respectives, donne en quelque sorte le taux « naturel » d'homogamie. Ainsi, l'absence d'association entre conjoints ne correspond pas à une homogamie nulle, qui signalerait au contraire une répulsion pour les individus partageant les mêmes caractéristiques que soi (homophobie ou hétérophilie, au sens littéral) : elle correspond à une homogamie égale à celle qu'on observerait en situation aléatoire.

L'évolution au fil des enquêtes du taux d'homogamie « naturel » ou attendu provoque donc une variation de l'homogamie observée même si les comportements de choix du conjoint, même si la tendance à choisir un individu dans le même groupe social que soi n'ont, eux, pas évolué. Qu'est-ce qui détermine la variation de ce niveau attendu ? La proximité des effectifs des hommes et des femmes au sein de chaque catégorie, et la taille relative des différentes catégories. En effet, pour que les proportions situées sur la diagonale de la table d'homogamie attendue soient élevées, il faut que les marges du côté des hommes et des femmes dont on fait le produit soient elles aussi élevées *en même temps*, c'est-à-dire pour la même catégorie des deux côtés. Ainsi, lorsque pour chaque catégorie les marges de la table d'homogamie sont très proches pour les hommes et pour les femmes, les catégories nombreuses vont avec les catégories nombreuses, et la somme des produits situés sur la diagonale sera plus élevée. Cet effet sera d'autant plus fort que certains groupes sont fortement sur-représentés sur les marges au détriment des autres : si la condition de symétrie hommes-femmes est respectée, le produit correspondant à la diagonale sera alors maximal[12].

10 Formellement, ces deux bornes se définissent par $Homo\ min = \sum_i Max(n_i^H + n_i^F - N,\ 0)$ et $Homo\ max = \sum_i Min(n_i^H, n_i^F)$, avec N, n_i^H, n_i^F respectivement le nombre total d'individus, le nombre d'hommes et de femmes de la catégorie i. Donnant un nombre d'individus, ces deux formules peuvent être rapportées à N pour obtenir un taux. *Cf.* Forsé et Chauvel (1995, pp. 127–128) pour plus de détails.

11 Limite rencontrée notamment par l'indice d'immobilité de Boudon, qui vaut schématiquement $H_B = \frac{Homo\ observée - Homo\ min}{Homo\ max - Homo\ min}$. *Cf.* encore une fois Forsé et Chauvel (1995, pp. 127–128) pour plus de détails.

12 Plus formellement, ces deux conditions (symétrie et concentration) découlent du fait que la somme de produits effectués deux à deux entre deux séries de proportions dont la somme est fixée pour chacune à 1 sera plus importante si l'on multiplie entre elles des proportions élevées que si l'on associe proportions élevées et proportions faibles (le maximum, 1, étant atteint pour la multiplication de deux proportions égales à 1, les autres étant par contrecoup à 0).

La concentration de la population

En %	1969	1979	1989	1999	2009
Diplôme (détaillé)	53,0	46,1	33,5	21,5	17,5
Diplôme (agrégé)	43,7	36,6	25,9	27,1	28,2
Classe Casmin	20,9	19,8	24,8	28,8	30,2
PCS	-	-	27,1	28,0	26,4

Tableau 6 : Concentration de la structure de la population

Champ : individus âgés de 30 à 59 ans à la date de l'enquête (célibataires inclus) ; pour la classe Casmin, seuls les actifs sont retenus.

Calcul : pour chaque année, avec p_i *la proportion d'individus de la catégorie i, et n le nombre de catégories, le taux de dissimilarité est égal à* $\frac{1}{2}\sum_i |p_i - \frac{1}{n}|$ *, soit la moitié de la somme des écarts des proportions à la valeur qui assurerait une distribution uniforme.*

Qu'en est-il concernant nos données ? Du point de vue de la concentration des marges sur quelques catégories (voir l'indice synthétique, tableau 6), l'évolution est assez nette pour l'éducation (Tableau 7) : alors qu'en 1969 les groupes des non diplômés et des titulaires du seul certificat d'études concentrent à eux deux 72 % des individus âgés de 30 à 59 ans, ils n'en totalisent plus que 22 % en 2009 ; à l'inverse, tous les autres niveaux voient leurs effectifs progresser, ce qui entraîne une nette réduction de la concentration, et donc de l'homogamie attendue[13]. L'effet est beaucoup moins net en ce qui concerne le statut socio-professionnel : la distribution selon la classification Casmin (Tableau 8) s'est concentrée d'une dizaine de points de pourcentage, alors qu'elle est restée stable selon la nomenclature des PCS (Tableau 9). Au total, on peut donc s'attendre à une forte baisse de l'homogamie attendue d'éducation, à une légère augmentation de celle selon la classe Casmin, et à une stagnation de celle concernant les PCS. Mais ces mouvements doivent compter avec les évolutions de l'asymétrie hommes femmes.

En %	Aucun	CEP	BEPC	CAP/ BEP	CAP/ BEP +BEPC	Bac pro/tech /BP/BT	Bac gén	1er cycle	2e et 3e cycles	École
1969	37,6	34,7	5,6	10,8	1,9	1,2	4,3	1,7	1,5	0,7
1979	29,6	29,5	5,9	17,0	3,4	1,4	5,7	3,6	3,1	0,9
1989	23,1	21,2	7,6	19,2	5,3	4,3	5,5	4,6	7,1	2,1
1999	20,4	11,1	7,5	19,6	10,3	4,9	6,5	9,7	7,2	2,7
2009	17,2	5,0	7,8	15,6	10,8	8,5	7,7	12,6	11,3	3,5

Tableau 7 : Structure par diplôme de la population

Champ : individus âgés de 30 à 59 ans à la date de l'enquête (célibataires inclus).

13 Cette conclusion rejoint celle de Goux et Maurin (2003, p. 62).

La nomenclature de classe Casmin

Aussi désignée sous le nom de schéma Erikson-Goldthorpe-Portocarero, la classification socio-économique Casmin (pour *Comparative Analysis of Social Mobility in Industrial Nations*) est une nomenclature internationale issue d'un projet initié par John Goldthorpe à la suite de travaux menés dans les années 1970 en Grande-Bretagne. Elle est depuis devenue la classification de référence des études comparatives, notamment dans le domaine de la mobilité sociale. Fondée principalement sur des critères économiques (source et niveau de revenu) et institutionnels (position du poste occupé au sein de la hiérarchie, autonomie), elle distingue 11 catégories :

- I : Professions libérales, cadres et responsables de niveau supérieur ; cadres dirigeants des grandes entreprises ; grands propriétaires.
- II : Professions libérales, cadres et responsables de niveau inférieur ; techniciens supérieurs ; cadres dirigeants des petites entreprises ; personnel d'encadrement (à l'exception des contremaîtres).
- IIIa : Employés de niveau supérieur (administrations).
- IIIb : Employés de niveau inférieur (commerces et services).
- IVa : Indépendants avec salariés (artisans, commerçants, chefs de petites entreprises...).
- IVb : Indépendants sans salariés (artisans, commerçants, chefs de petites entreprises...).
- IVc : Agriculteurs exploitants et petits propriétaires ; indépendants du secteur primaire.
- V : Techniciens de niveau inférieur ; contremaîtres.
- VI : Ouvriers qualifiés.
- VIIa : Ouvriers non-qualifiés hors agriculture.
- VIIb : Ouvriers agricoles et autres ouvriers du secteur primaire.

En %	I	II	IIIa	IIIb	IVa	IVb	IVc	V	VI	VIIa	VIIb
1969	6,4	7,9	10,4	8,3	3,3	9,3	11,7	5,1	14,5	20,5	2,6
1979	9,2	10,5	12,0	10,1	2,8	6,9	8,8	5,7	14,9	17,6	1,4
1989	11,0	13,8	14,8	11,3	3,0	5,4	5,6	5,5	14,8	13,7	1,0
1999	12,8	13,5	16,6	13,2	3,2	4,2	3,0	5,2	15,0	12,3	1,1
2009	16,4	15,5	15,1	14,6	0,7	6,1	1,8	5,8	12,6	10,7	0,8

Tableau 8 : Structure par classe Casmin de la population

Champ : actifs âgés de 30 à 59 ans à la date de l'enquête (célibataires inclus).

En %	Agriculteurs	Indépendants	Cadres	Intermédiaires	Employés	Ouvriers
1982	6,8	8,4	8,4	18,4	26,8	31,2
1989	5,2	7,9	9,8	19,5	28,2	29,4
1999	2,9	7,0	12,1	19,9	30,8	27,3
2009	1,7	6,4	15,5	22,5	30,6	23,4

Tableau 9 : Structure par PCS de la population

Champ : individus âgés de 30 à 59 ans à la date de l'enquête (célibataires inclus).

L'équilibre entre hommes et femmes

On l'a dit, le taux brut d'homogamie est aussi déterminé de manière importante par l'équilibre entre proportion d'hommes et de femmes au sein de chacune des catégories utilisées. Comment a évolué cet équilibre, synthétisé par un indice de similarité entre les populations des hommes et des femmes (Tableau 10) ? En ce qui concerne l'éducation, tout d'abord, l'élévation du niveau moyen, et particulièrement de celui des femmes, n'a pas influencé cette similarité dans un sens ou dans un autre[14] : quelle que soit la classification utilisée, les mouvements de l'indice sont faibles et irréguliers. Il en va de même pour la structure socio-professionnelle de la population mesurée en termes de PCS. En revanche, on observe un accroissement de 6 points de la dissimilarité entre hommes et femmes du point de vue de la classe Casmin, ce qui laisse anticiper une baisse mécanique de l'homogamie selon ce critère ; cet effet va dans un sens opposé à celui dû à l'augmentation de la concentration de la population : le résultat, indéterminé, risque d'être relativement stable, les deux mouvements étant d'ampleur comparable.

Au niveau détaillé (tableau 11), on peut vérifier que la féminisation des diplômes les plus élevés (grandes écoles, second cycle universitaire et baccalauréat général, le premier cycle étant déjà très féminisé dès 1969) a mené selon les cas aussi bien à plus de dissimilarité hommes-femmes (cas du baccalauréat général) qu'à plus de similarité (cas des grandes écoles), ou à une dissimilarité plus faible, mais inversée (cas du second cycle universitaire, où les femmes sont devenues majoritaires). Au niveaux inférieurs, les mouvements sont divers : féminisation menant à moins de similarité dans le cas du certificat d'études, à plus de similarité dans ceux des non diplômés et des titulaires du CAP/BEP, tendances irrégulières ailleurs. On comprend donc comment, au niveau global, la dissimilarité en termes de diplôme reste sensiblement stable.

En ce qui concerne la classe Casmin (tableau 12), la réduction de la dissimilarité hommes-femmes provient notamment de la féminisation des groupes situés aux extrêmes : les classes supérieures (I) d'un côté, les ouvriers agricoles de l'autre (VIIb). Mais elle est compensée par l'accroissement de la dissimilarité qui provient de la féminisation de professions au sein desquelles les femmes étaient déjà assez présentes ou sur-représentées : les professions intermédiaires et les employés (II, IIIa et IIIb) ; mais aussi de la masculinisation du groupe des indépendants sans salariés (IVb) et des agriculteurs (IVc). On retrouve en fait ces évolutions en des termes plus familiers, et sur une période plus courte, à travers les PCS (tableau 13).

En %	**1969**	**1979**	**1989**	**1999**	**2009**
Diplôme (détaillé)	6,5	6,9	7,6	7,7	6,5
Diplôme (agrégé)	6,0	5,7	5,6	5,4	5,6
Classe Casmin	14,2	15,2	18,5	20,4	20,6
PCS	-	-	17,7	18,5	18,0

Tableau 10 : Dissimilarité entre les niveaux d'études des hommes et des femmes

Champ : individus âgés de 30 à 59 ans à la date de l'enquête (célibataires inclus) ; pour la classe Casmin, seuls les actifs sont retenus.

Lecture : en 1969, 6,5 % des individus étaient forcés de choisir un conjoint de niveau d'études différent pour que l'ensemble de la population des cohortes étudiées se mette en couple au sein des mêmes cohortes.

14 Ce résultat est, encore une fois, en accord avec Goux et Maurin (2003, p. 62).

Calcul : pour chaque année, soit n_d^H (respectivement n_d^F) le nombre d'hommes (de femmes) dont le diplôme le plus élevé est d, et N le nombre total d'individus, le taux de dissimilarité est égal à $\frac{1}{2N}\sum_d |n_d^H - n_d^F|$ *. Cet indice, légèrement différent de l'indice de dissimilarité utilisé notamment en géographie, présente dans notre cas l'intérêt de comparer réellement le nombre d'hommes et de femmes d'une catégorie, au lieu de mesurer des écarts abstraits entre taux .*

En %	**Aucun**	**CEP**	**BEPC**	**CAP/ BEP**	**CAP/ BEP +BEPC**	**Bac pro/tech /BP/BT**	**Bac gén**	**1er cycle**	**2e et 3e cycle**	**École**
1969	-5,8	-9,6	-18,2	38,1	-9,5	58,5	-6,2	-19,5	31,9	77,7
1979	-3,6	-12,8	-22,5	27,9	-18,7	44,5	-1,0	-17,6	13,9	77,6
1989	-4,4	-15,0	-19,2	25,9	-14,1	14,3	-14,3	-18,1	-7,4	74,2
1999	-4,2	-19,1	-18,3	25,0	-8,9	1,9	-23,5	-18,4	-6,1	52,6
2009	-1,7	-14,8	-12,6	23,6	-4,6	15,0	-21,9	-19,5	-9,0	31,5

Tableau 11 : Évolution de l'équilibre hommes-femmes par diplôme

Champ : individus âgés de 30 à 59 ans à la date de l'enquête (célibataires inclus).

Lecture : en 2009, 23,6 % des titulaires d'un CAP/BEP devaient choisir un conjoint de niveau d'études différent s'ils souhaitaient se mettre en couple au sein des cohortes étudiées ; ces individus étaient des hommes.

Calcul : pour chaque année, soit n_d^H (respectivement n_d^F) le nombre d'hommes (de femmes) dont le diplôme le plus élevé est d, le taux de dissimilarité est égal à $\frac{n_d^H - n_d^F}{n_d^H + n_d^F}$ *.*

En %	**I**	**II**	**IIIa**	**IIIb**	**IVa**	**IVb**	**IVc**	**V**	**VI**	**VIIa**	**VIIb**
1969	67,4	0,0	-0,8	-48,5	70,3	-0,8	11,6	87,6	80,0	44,1	72,9
1979	53,3	-12,2	-19,7	-49,7	67,1	1,8	13,7	83,1	77,0	30,5	73,1
1989	46,0	-14,1	-46,6	-62,7	61,7	15,3	19,0	84,9	76,9	32,6	52,2
1999	31,6	-22,1	-54,2	-64,8	60,5	20,7	28,8	79,2	69,7	34,4	39,2
2009	22,0	-31,5	-53,4	-66,3	71,8	38,2	39,1	76,8	72,9	42,0	29,9

Tableau 12 : Évolution de l'équilibre hommes-femmes par classe Casmin

Cf. *Tableau 11 pour les détails.*

En %	**Agriculteurs**	**Indépendants**	**Cadres**	**Intermédiaires**	**Employés**	**Ouvriers**
1982	7,2	21,9	50,4	14,2	-61,8	36,3
1989	12,1	25,9	42,0	10,0	-60,7	38,1
1999	25,6	34,2	29,0	2,7	-62,1	44,4
2009	37,4	40,1	19,4	-4,1	-60,7	49,4

Tableau 13 : Évolution de l'équilibre hommes-femmes par PCS

Cf. *Tableau 11 pour les détails.*

Résultats

Les figures 1, 2 et 3 (pages suivantes) présentent les évolutions des trois mesures de l'homogamie auxquelles nous nous intéressons[15] ; sont représentées en même temps l'homogamie observée, correspondant à la part des couples qui se trouvent sur la diagonale de la table d'homogamie, et l'homogamie attendue en situation de mise en couple aléatoire. Il est tout d'abord intéressant de noter que les trois types d'homogamie baissent légèrement en termes bruts. La plus forte baisse concerne l'homogamie d'éducation, qui passe de 45 à 30 % en quarante ans, soit une diminution d'un tiers. L'homogamie socio-professionnelle baisse bien plus faiblement, perdant entre 5 et 10 points selon que l'on se réfère à la classe Casmin ou à la classification des PCS.

Mais ces variations brutes observées ne s'accompagnent pas d'évolutions similaires de l'homogamie attendue en situation aléatoire. La baisse mesurée de l'homogamie d'éducation découle pour une grande part d'un affaiblissement parallèle du niveau attendu sous hypothèse de choix aléatoire du conjoint, que nous avions prévu plus haut ; au contraire, la réduction observée de l'homogamie socio-professionnelle s'oppose à une légère augmentation de ce qui est attendu dans le cas de la classe Casmin, et à une baisse à peine perceptible de celle-ci dans le cas des PCS (deux évolutions qui correspondent aussi à ce qu'on pouvait attendre). Il est donc clair que l'évolution de l'homogamie d'éducation est due pour sa plus grande part, sinon dans son intégralité, à des variations dans la structure éducative de la population ; celle-ci s'est en effet très fortement transformée sur la période étudiée, alors que la structure socio-professionnelle est restée plus stable. Comme nous l'avons expliqué plus haut, le problème de la mesure brute de l'homogamie est qu'elle ne permet pas de savoir de manière rigoureuse si l'homogamie a véritablement changé, ou si ce sont simplement les catégories utilisées qui ne décrivent pas avec la même précision la population d'hier et d'aujourd'hui. Il est donc nécessaire de recourir à des mesures plus robustes.

15 La quatrième mesure, constituée par une classification plus agrégée des diplômes, n'est pas présentée ici puisqu'elle reproduit essentiellement le même type d'évolution que la classification détaillée, mais avec des taux légèrement plus élevés. Dans la suite, nous ne signalerons pas les résultats obtenus avec la classification agrégée lorsqu'ils n'apportent rien de plus.

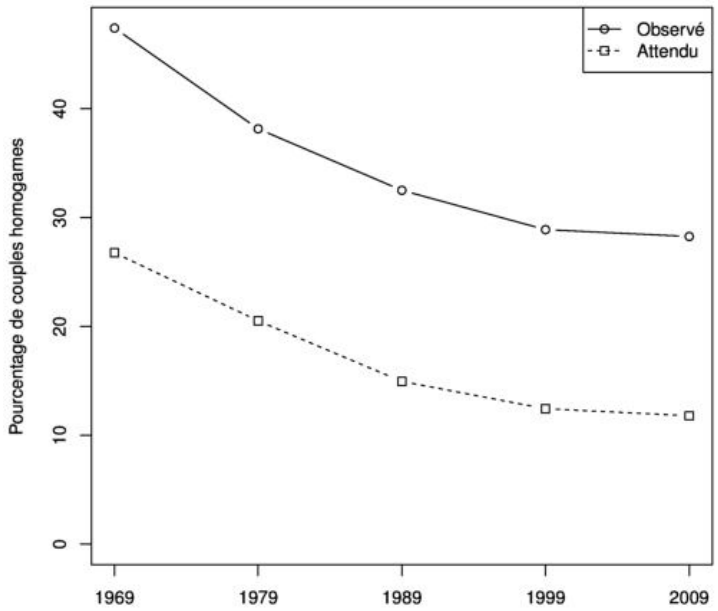

Figure 1 : Évolution de l'homogamie brute d'éducation, nomenclature détaillée

Champ : couples cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

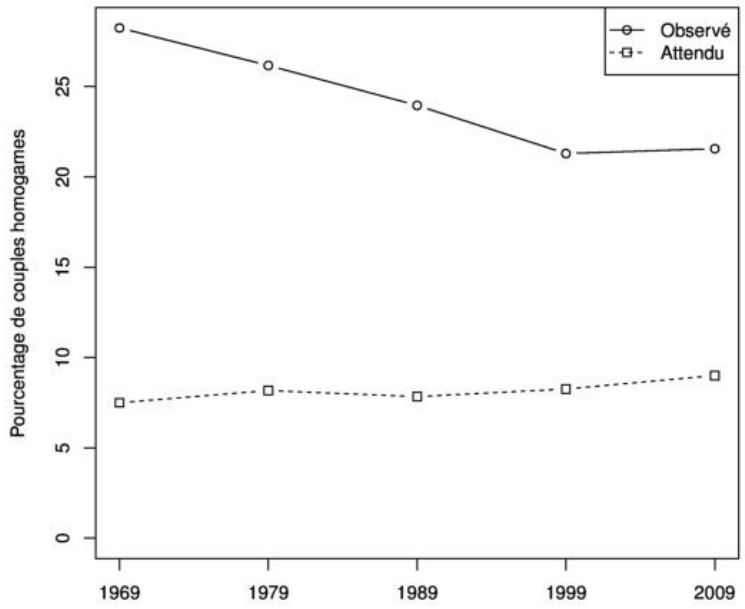

Figure 2 : Évolution de l'homogamie brute selon la classe Casmin

Champ : couples cohabitants d'actifs dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

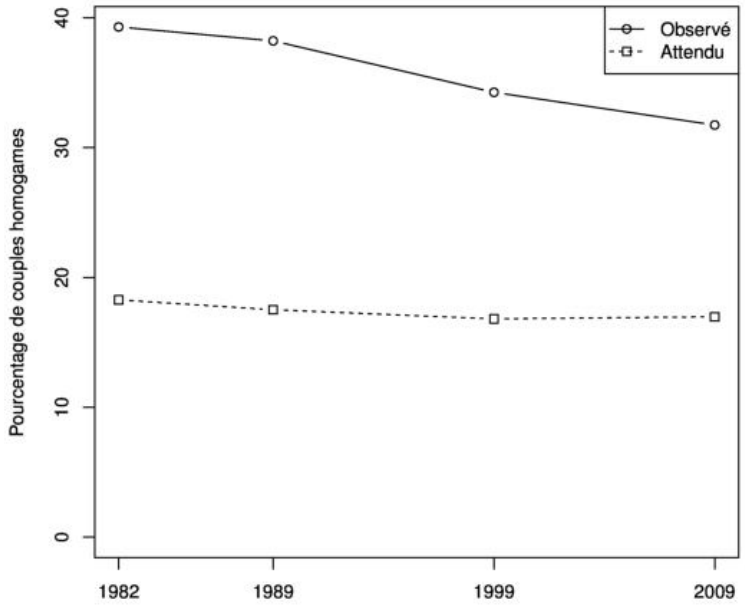

Figure 3 : Évolution de l'homogamie brute selon la PCS

Champ : couples cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

III L'homogamie nette : un affaiblissement réel

Parce qu'elle est beaucoup plus résistante à la variation des marges et à la redéfinition des catégories, l'homogamie nette est la mesure qui présente de loin le plus d'intérêt. Le principe de cette approche est de recourir, de près ou de loin, à l'outil statistique des *odds ratios*[16] : décrivant l'association intrinsèque entre les caractéristiques des conjoints dans une table d'homogamie, cet indicateur donne accès non plus simplement aux proportions de couples observées, mais à la tendance à choisir un conjoint d'un type donné, indépendamment de l'effet dû à la structure de la population. Ainsi, l'*odds ratio* permet non seulement d'éliminer, lors des comparaisons entre enquêtes, les variations dues aux transformations intervenues dans la composition de la population ; il nous prémunit aussi contre les artefacts liés à la construction des catégories que nous avons évoqués plus haut[17].

Sur une table d'homogamie, comme sur une table de mobilité, plusieurs types d'*odds ratios* peuvent être calculés, correspondant chacun un type de couple. Si l'on s'attache à décrire l'homogamie au sens strict, soit la tendance à choisir un conjoint exactement dans le même groupe social, qui correspond à la diagonale de la table, l'*odds ratio* le plus simple mesure, précisément, les chances pour un individu de la catégorie considérée de se mettre en couple au sein de cette catégorie plutôt qu'en dehors, rapportées à celles d'un individu non membre de la catégorie en question de se mettre en couple au sein de la catégorie plutôt qu'en dehors. Ceci s'exprime plus directement en termes mathématiques ; si l'on se fonde sur le tableau suivant, exprimé en proportions ou en effectifs :

Homme	**Femme**	
	Catégorie *i*	**Catégorie *j***
Catégorie *i*	n_{ii}	n_{ij}
Catégorie *j*	n_{ji}	n_{jj}

La formule de l'*odds ratio* en question est alors : $od_{ii} = \frac{n_{ii}/n_{ij}}{n_{ji}/n_{jj}} = \frac{n_{ii}n_{jj}}{n_{ij}n_{ji}} = \frac{n_{jj}/n_{ji}}{n_{ij}/n_{ii}} = od_{jj}$

Lorsque l'on a affaire à une table de plus de deux catégories, ce qui est généralement le cas, le principe reste le même : il suffit de ramener, pour le calcul de chaque *odds ratio* individuel, la table au format ci-dessus, en faisant la somme de toutes les lignes et colonnes correspondant aux catégories autres que celle étudiée.

16 Pour une présentation de l'histoire, de l'intérêt et des propriétés de l'*odds ratio*, *cf.* Vallet (2007).

17 Il serait cependant abusif, et même lâche, de la part du sociologue, de s'en remettre intégralement à l'outil statistique pour résoudre les problèmes théoriques qui se posent à lui : si l'*odds ratio* donne bien une mesure robuste de l'association entre groupes, il ne résout pas par magie la question de la définition des groupes. Il nous dira par exemple quelle est la tendance à la formation de couples ouvrier-ouvrière, employé-employée, ou encore ouvrier-employée de manière fiable, mais ne suggèrera pas au chercheur peu attentif qu'il existe aussi une homogamie propre au groupe formé par la fusion des PCS ouvriers et employés, plus forte que les tendances respectives à l'homogamie des groupes pris séparément, à cause de la forte association entre les deux groupes (en 2009, les *odds ratios* mesurant l'homogamie sont respectivement de 1,8 pour les employés, 4,0 pour les ouvriers, et 4,9 pour les deux groupes pris ensemble).

Plus généralement, l'*odds ratio* peut être utilisé pour mesurer l'association entre deux catégories différentes du côté de l'homme et de la femme ; il fonctionne alors exactement de la même manière. En reprenant le tableau ci-dessus, il est possible de calculer les chances de voir un certain type de couple hétérogame se former :

$$od_{ij} = \frac{n_{ij}/n_{ii}}{n_{jj}/n_{ji}} = \frac{n_{ij}n_{ji}}{n_{jj}n_{ii}} = \frac{n_{ji}/n_{jj}}{n_{ii}/n_{ij}} = od_{ji} = \frac{1}{od_{ii}} = \frac{1}{od_{jj}}$$

Il est intéressant de noter qu'une table à quatre cases est en fait décrite par un seul *odds ratio*, qui mesure les chances de former un couple homogame pour les deux catégories aussi bien que l'inverse des chances de ne pas le faire, les quatre valeurs étant conceptuellement et mathématiquement liées.

L'indice de diagonalité nette

Plusieurs voies permettent de passer de la mesure détaillée décrivant l'association entre deux caractéristiques données des conjoints (par exemple, les PCS de l'homme et de la femme) à une mesure globale de la force de l'association. Avant de recourir à des modèles log-linéaires et log-multiplicatifs plus complexes, nous utilisons ici un indicateur qui n'a pas connu un grand succès, probablement à cause de l'existence des premiers, mais qui présente l'intérêt d'être relativement simple : l'indice de diagonalité nette proposé par Forsé et Chauvel (1995). Son calcul nous permettra notamment de confronter nos résultats à ceux exposés dans leur article.

Cet indice est construit à partir du produit de tous les *odds ratios* individuels mesurant les chances de choisir un conjoint dans la même catégorie que soi plutôt que dans chacune des autres catégories. Il prend la valeur 0 en situation aléatoire, lorsque tous les *odds ratios* sont égaux à 1, et tend vers 1 lorsque l'homogamie se rapproche de la perfection (le produit des *odds ratios* tend alors vers l'infini). Cet indice mesure donc la force de l'homogamie au sens strict (choix d'un conjoint exactement dans la même catégorie), correspondant à la diagonale de la table, mais corrige les défauts de la mesure brute de l'homogamie. Sa formule est :

$$IDN = \frac{\ln \prod_{i<j} \frac{n_{ii}n_{jj}}{n_{ij}n_{ji}}}{\frac{k(k-1)}{2} + \ln \prod_{i<j} \frac{n_{ii}n_{jj}}{n_{ij}n_{ji}}}$$, avec $\prod_{i<j} \frac{n_{ii}n_{jj}}{n_{ij}n_{ji}}$ le produit de tous les *odds ratios* qui peuvent être construits sur la diagonale de la table, et k le nombre total de catégories.

Quelle évolution décrit cet indice ? La figure 4 représente la valeur de l'indice à chaque enquête pour les trois dimensions étudiées. L'homogamie d'éducation, qui était la plus faible en 1969, a décliné selon les trois nomenclatures, mais à des rythmes très contrastés. Si l'on suit la nomenclature détaillée ou celle de Forsé et Chauvel en quatre groupes, elle a diminué jusqu'en 1989, avant de se stabiliser, voire d'augmenter à nouveau[18]. En revanche, si l'on suit la classification agrégée, intermédiaire entre les deux précédentes, on observe une chute bien plus rapide, ainsi qu'une remontée plus tardive. Cette relative divergence des résultats selon la classification utilisée a de quoi surprendre, surtout par le fait que c'est la classification

18 Étant donné que nous n'avons pas calculé d'intervalles de confiance pour l'indice de diagonalité nette, il est risqué d'interpréter les faibles évolutions observées pour l'éducation après 1989. On se réfèrera aux modèles log-multiplicatifs présentés plus bas pour une estimation de ces intervalles et de la significativité de ces mouvements.

intermédiaire qui fait apparaître une évolution invisible à un niveau plus détaillé ou plus agrégé. Nous touchons probablement là une limite de l'indice de diagonalité nette, qui reste assez sensible à la définition des catégories, puisqu'il mesure l'homogamie stricte (correspondant à la diagonale de la table), qui dépend des regroupements effectués ; nous verrons en effet plus loin qu'il est possible de corriger cet indice par des pondérations, et surtout que les modèles log-multiplicatifs sont plus robustes. Un élément intéressant, sur lequel les trois classifications convergent, tient à la légère remontée, ou du moins à la stabilisation qui s'observe entre 1999 et 2009 ; nous y reviendrons.

Les résultats concernant l'homogamie socio-professionnelle, obtenus à l'aide de deux classifications très différentes, convergent en revanche. Ce type d'homogamie a décru de manière relativement régulière depuis 1969 de 0,03 point d'indice, baisse qui se situe donc entre les deux évolutions de l'homogamie d'éducation observées avec des classifications différentes. Notons que ces mouvements sont d'une ampleur limitée, puisque 0,03 point de décrue correspond à un peu moins de 4 % de la valeur de l'indice. Pour faible qu'il apparaisse, le mouvement de décrue semble cependant assez régulier pour que l'on soit assuré de sa réalité. De plus, le rythme ne semble pas avoir ralenti au cours des dernières enquêtes, contrairement à celui de l'homogamie d'éducation.

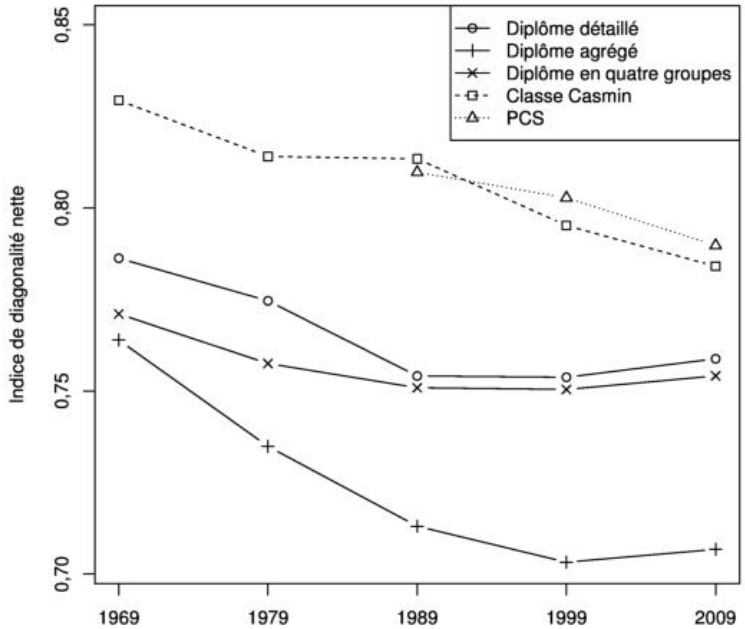

Figure 4 : Évolution de l'indice de diagonalité nette

Champ : couples cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête (restreints aux actifs dans le cas de la classe Casmin).

Nos résultats peuvent-ils être comparés avec ceux de Forsé et Chauvel ? Les deux générations étudiées par les auteurs à partir de la seule enquête Emploi 1989 sont formées des cohortes nées avant 1931, pour la plus ancienne, et de celles nées après 1953, pour la plus récente (à la différence de notre échantillon, c'est l'âge de l'épouse qui compte dans leur sélection). Si l'on se réfère au tableau 5 (page 40), la première est donc présente dans

l'échantillon *grosso modo* en 1969 et 1979, alors que la seconde fait son apparition en 1989, au moment justement où la précédente disparaît. L'évolution qu'ils notent doit donc prendre place progressivement, dans nos données, entre les enquêtes de 1969 et 1979, et celle de 1989. Il n'est pas possible de comparer précisément les niveaux de l'indice, à cause de la présence des autres cohortes ; c'est donc l'existence d'une variation et son ampleur qui nous intéressent.

Concernant l'éducation, les auteurs observent une stabilité de l'homogamie, avec un indice de 0,75[19]. Si l'ordre de grandeur est bien le même, nous ne trouvons clairement pas la même variation : en utilisant notre classification détaillée des diplômes, l'affaiblissement de l'homogamie d'éducation a lieu intégralement sur les trois premières enquêtes, qui décrivent justement les cohortes étudiées par Forsé et Chauvel ; si nous utilisons la classification des auteurs en quatre groupes[20], nous observons une réduction un peu moins marquée, mais cependant visible ; enfin, si nous utilisons la classification intermédiaire, les résultats diffèrent du tout au tout. Nous ne pouvons avancer qu'une explication à cette divergence de résultats : l'utilisation d'une seule enquête au lieu de plusieurs. Puisque nous avons pu reproduire les résultats de Forsé et Chauvel à partir de l'enquête Emploi 1989 en conservant la même définition d'échantillon et les mêmes classifications de diplôme, la seule différence entre les deux approches tient à l'existence du biais d'attrition. Même s'il convient de rester prudent dans l'attente de plus amples vérifications, il est donc possible que l'écart entre nos résultats et les leurs tienne au fait que les couples hétérogames du point de vue de l'éducation présentent une durée de vie plus courte. Cela signifierait alors que l'intégralité de l'affaiblissement de l'homogamie d'éducation intervenu avant 1989 provient de l'effet des séparations et des remises (ou non) en couple, hypothèse qui semble de prime abord assez étonnante, même si les évolutions observées en utilisant la classification des auteurs sont très faibles.

En revanche, nos résultats sont mieux en accord concernant les deux autres dimensions de l'homogamie. Forsé et Chauvel observent un affaiblissement de l'homogamie d'origine sociale (passage de 0,73 à 0,70 entre les deux générations étudiées) qui s'approche de ce que nous observons pour l'homogamie socioprofessionnelle. Bien qu'il soit difficile d'aller plus loin dans la comparaison, ces deux résultats semblent converger dans le sens d'une réduction de la détermination socio-professionnelle, qu'elle soit héritée ou directe, du choix du conjoint. Celle-ci reste cependant, dans nos résultats, plus importante que celle liée à l'éducation ; c'était l'inverse chez Forsé et Chauvel, mais cela n'est pas étonnant, puisque l'origine sociale est probablement moins déterminante que les caractéristiques individuelles comme l'éducation et la profession.

Tenir compte de la nomenclature

L'indice de diagonalité nette tel que proposé par Forsé et Chauvel peut être amélioré de manière à tenir compte de la répartition des effectifs dans les différentes catégories utilisées par les nomenclatures. En effet, en l'état, il donne un poids équivalent à toutes les catégories, ce qui a relativement peu d'importance dans leur cas, puisque leurs catégories sont de tailles équivalentes, mais qui peut expliquer pourquoi notre nomenclature intermédiaire diverge en partie des autres.

---

19 Résultat que nous retrouvons effectivement si nous calculons l'indice de diagonalité nette par cohortes décennales (selon l'âge de l'homme) à partir de l'enquête Emploi 1989 : il est égal à 0,76 pour toutes les cohortes, soit une valeur sensiblement identique étant donné que nous incluons aussi les couples non mariés. Les autres classifications amènent aussi à conclure à une stabilité.

20 *Cf.* Tableau 2, page 37.

La formule de l'indice de diagonalité nette correspond en fait à la moyenne géométrique des *odds ratios*. En effet (Forsé et Chauvel, 1995, n. 3) :

$$IDN = \frac{\ln\theta}{\frac{k(k-1)}{2}+\ln\theta} = \frac{\ln\theta^{\frac{2}{k(k-1)}}}{1+\ln\theta^{\frac{2}{k(k-1)}}} \quad \text{avec} \quad \ln\theta^{\frac{2}{k(k-1)}} = \ln \sqrt[\frac{k(k-1)}{2}]{\prod_{i<j}\frac{n_{ii}n_{jj}}{n_{ij}n_{ji}}}$$

le logarithme naturel de la moyenne géométrique non pondérée des $\frac{n_{ii}n_{jj}}{n_{ij}n_{ji}}$ (ce que ne précisent pas les auteurs), puisque $\frac{k(k-1)}{2}$ correspond au nombre d'*odds ratios* calculables à partir de la diagonale d'une table carrée de dimension k (en fait, au nombre de cellules situées dans un triangle de la table, en excluant la diagonale).

Partant de ce principe, il est assez direct d'aménager l'indice de diagonalité nette de manière à inclure des pondérations : nous n'avons qu'à modifier θ afin d'utiliser une moyenne géométrique pondérée, en prenant comme coefficient de pondération les effectifs inclus dans les quatre cases utilisées pour calculer chaque *odds ratio*, soit $w_{ij}=n_{ii}+n_{jj}+n_{ij}+n_{ji}$. On obtient :

$$\ln\theta = \ln \sqrt[\sum_{i<j} w_{ij}]{\prod_{i<j}\left(\frac{n_{ii}n_{jj}}{n_{ij}n_{ji}}^{w_{ij}}\right)} = \ln\left[\prod_{i<j}\left(\frac{n_{ii}n_{jj}}{n_{ij}n_{ji}}\right)^{w_{ij}}\right]^{\frac{1}{\sum_{i<j} w_{ij}}} = \frac{\sum_{i<j} w_{ij}\ln\frac{n_{ii}n_{jj}}{n_{ij}n_{ji}}}{\sum_{i<j} w_{ij}}$$

La version révisée de l'indice de diagonalité nette qui découle de cette redéfinition de θ possède les mêmes propriétés que l'original, mais donne à certains *odds ratios* plus d'importance qu'à d'autres. Elle est donc plus juste de manière générale, et permet en particulier dans notre cas de comparer plusieurs classifications.

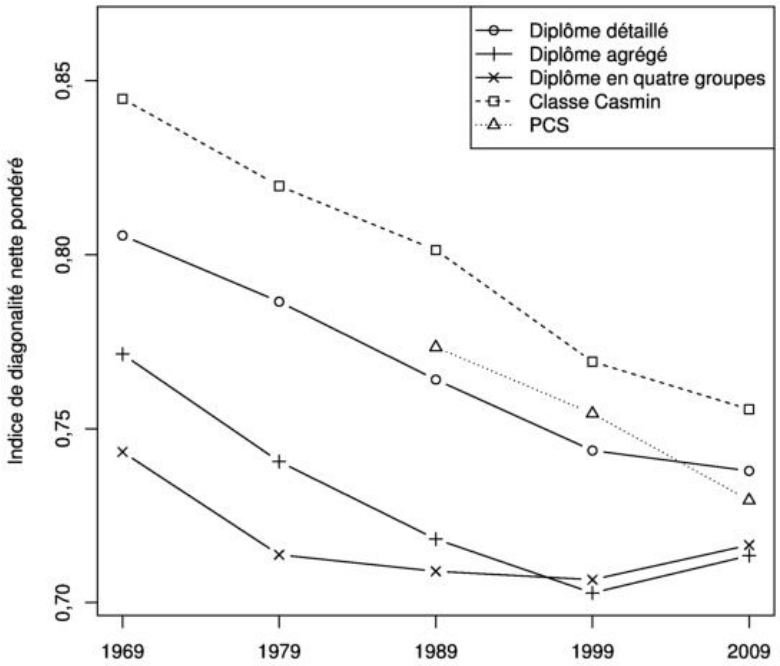

Figure 5 : Évolution de l'indice de diagonalité nette pondéré

Champ : couples cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête (restreints aux actifs dans le cas de la classe Casmin).

Avec cette nouvelle version de l'indice, toutes les classifications du diplôme mènent à la même conclusion (Figure 5) : l'homogamie d'éducation s'est réduite en quarante ans, même si les rythmes diffèrent, ce qui est attendu (notamment parce qu'on ne tient compte ici que de l'homogamie au sens strict). En fait, notre indice corrige autant la classification utilisée par les auteurs, en augmentant sa pente, que notre classification intermédiaire, en la réduisant. Avec notre indice, les auteurs auraient conclu, à partir de leurs données, à un très léger affaiblissement de l'homogamie d'éducation : dans la seule enquête Emploi 1989, cet indice passe de 0,730 pour les 50-59 ans à 0,727 pour les 40-49 ans, puis 0,721 pour les 20-29 ans (la décroissance est la même avec les autres classifications).

Il n'est pas étonnant que les classifications détaillées obtiennent des valeurs d'indice plus élevées : une fois tenu compte du nombre de catégories utilisées, l'homogamie se renforce à mesure que celles-ci deviennent précises, puisque le degré de proximité entre individus augmente. Ce mécanisme n'est pas une loi absolue : il est possible d'imaginer un niveau de détail au-delà duquel les distinctions introduites, trop subtiles, ne décrivent plus des proximités sociales réellement significatives. Tant qu'il est observé, cet effet garantit en revanche l'intérêt de l'indice pondéré ainsi que la validité des classifications : c'est justement l'objectif de la mesure de l'homogamie nette que de mettre de côté les artefacts dus à la mesure elle-même.

Enfin, il apparaît toujours que sur la période récente, l'homogamie d'éducation s'est stabilisée ou a commencé à se renforcer. Ce phénomène, qui était déjà visible avec la version originale de l'indice, se trouve partiellement confirmé par une autre méthode, celle des modèles log-linéaires et log-multiplicatifs, pour lesquels nous avons cette fois calculé des indicateurs de significativité statistique.

Modèles log-linéaires et log-multiplicatifs

Si l'indice de diagonalité nette présente l'avantage d'une certaine simplicité, et surtout de fournir une vision synthétique des différents types d'homogamie, de leurs évolutions et de leurs niveaux, on peut lui adresser plusieurs reproches. Tout d'abord, il ne décrit que l'association observée sur la diagonale de la table, autrement dit l'homogamie stricte, et néglige complètement la question plus large des distances entre groupes sociaux. Ensuite, même s'il est possible de le coupler avec des tests de significativité statistique, il n'est pas possible de confronter l'hypothèse qu'il implique de faire, selon laquelle la diagonale de la table possède une réalité sociologique et une unité de comportement, à d'autres interprétations. Enfin, cet indice n'a jamais été réutilisé dans la littérature : pour rendre nos résultats comparables à ceux des travaux existants, il est nécessaire de recourir à d'autres approches.

Les modèles log-linéaires et log-multiplicatifs répondent à ces trois critiques. Comme on l'a vu dans la première partie, certains types de modèles permettent de tenir compte de l'homogamie au sens large, en faisant des hypothèses que l'on peut vérifier quant à la structuration de l'espace social et à ses transformations. Outil privilégié de l'étude de l'homogamie – comme ils le sont pour celle de la mobilité sociale –, ils autorisent par ailleurs des comparaisons assez directes avec la littérature existante. Nous avons donc eu recours à une large palette de modèles dans le but de savoir quelle est la meilleure manière de décrire l'homogamie et son évolution.

Nous ne détaillerons pas ici l'ensemble des modèles testés, mais en résumerons simplement le principe : on se réfèrera pour plus de détails à la présentation des diverses familles de modèles faite en première partie (« Les familles de modèles log-linéaires et log-multiplicatifs », page 12). L'approche retenue ici consiste à croiser les différents types possibles de modélisation de l'association entre caractéristiques des conjoints avec différentes modalités de variation temporelle (Tableau 14). On part de la description parfaite de l'association (association complète), puis on simplifie cette description en supposant successivement une symétrie des associations homme-femme (quasi-symétrie), un classement des catégories sur une seule dimension identique pour les hommes et les femmes (association lignes-colonnes homogène), une association variant avec le nombre de catégories à franchir (distance) ou avec le niveau des catégories traversées (franchissement)[21]. Ces manières de représenter l'espace social peuvent chacune s'associer avec plusieurs modélisations de la transformation observée au fil des enquêtes : stabilité (aucune variation), variation libre de chaque association à chaque enquête, variation uniforme de toutes les associations à chaque enquête, variation uniforme linéaire de toutes les associations en fonction du nombre d'années

21 Les modèles de type distance et franchissement n'ayant sens que pour décrire des catégories ordonnées hiérarchiquement sur une seule dimension, nous les avons classées en fonction des scores qu'elles obtiennent dans un modèle lignes-colonnes à une dimension. Pour l'échelle des diplômes, il suit l'ordre de la classification officielle, à part en ce qui concerne le BEPC, qui se place toujours plus haut que le CAP/BEP. Pour les PCS, ce classement est relativement original : Ouvriers, Agriculteurs, Employés, Indépendants, Professions intermédiaires, Cadres et professions intellectuelles supérieures. Pour Casmin : VIIa, VIIb, VI, IIIb, IVc, V, IIIa, IVb, II, IVa, I.

écoulées entre enquêtes. Finalement, ces variations peuvent affecter soit l'ensemble des associations, soit la seule diagonale, soit encore la diagonale et le reste de la table, mais selon des dynamiques différentes. Au total, on teste donc cinq familles de modèles, avec neuf variations différentes appartenant à quatre grands groupes[22].

Pour toutes les dimensions de l'homogamie, on peut constater que les variations observées restent d'ampleur relativement faible, puisque le modèle supposant une stabilité de l'association (spécifiée de manière complète mais sans variation temporelle) est déjà de bonne qualité : la part d'individus mal classés n'est que de 3,88 % pour l'éducation, de 4,88 % pour la classe Casmin, et de 1,87 % pour la PCS, et doit être préféré au modèle saturé selon l'indicateur BIC[23] (on se réfèrera aux tableaux 14, 17 et 18 ci-dessous pour les détails). Ce résultat est à relier aux remarques déjà faites à propos de l'homogamie brute : le modèle de stabilité tient compte des variations structurelles dans la composition de la population, qui sont à l'origine de la majorité de l'évolution de l'homogamie observée et attendue en situation de choix aléatoire du conjoint ; la part d'individus qui ne sont pas placés correctement par le modèle correspond donc à la divergence entre l'homogamie observée et celle qui serait attendue en situation de hasard.

Le fait que cette part soit relativement faible ne doit cependant pas nous tromper : même si la variation temporelle affecte peu d'individus, elle n'en est pas moins le véritable indicateur de la force de l'homogamie, une fois déduite l'influence mécanique qu'exerce la composition de la population sur les probabilités (et les nécessités, sous peine de rester célibataire) de choix de tel conjoint plutôt que de tel autre. Dans une logique cherchant à mesurer les comportements de choix du conjoint, signe de l'ouverture d'une société et de l'importance donnée à certaines caractéristiques sociales, plus que le résultat que constitue le nombre de couples de chaque type, c'est bien cette association nette qui présente un véritable intérêt, aussi faible que soit sa variation en termes d'effectifs.

Si nous cherchons à mesurer l'ampleur de la variation observée en termes d'homogamie nette, deux cas de figure se présentent. D'une part, si les modèles postulant une variation libre de l'homogamie apparaissent comme les meilleurs, nous devrons conclure que l'association entre conjoints a connu des variations désordonnées, autrement dit que la structure de l'association s'est transformée, sans préjuger du niveau de cette transformation. En revanche, si les modèles de variation uniforme (linéaire ou libre) sont de meilleure qualité que les précédents, cela signifie que la structure de l'association est restée la même, et qu'elle a simplement changé de niveau, niveau qui nous sera indiqué par le paramètre de variation. En fait, quelle que soit la dimension étudiée, ce sont toujours les variations uniformes qui

22 D'autres familles de modèles auraient pu être testées, notamment des associations lignes-colonnes à plusieurs dimensions, auxquelles nous recourons plus bas (*cf.* « Les transformations de l'espace social », page 68). Cependant, le paquet gnm que nous utilisons ne prend pas en charge la combinaison de ces types d'association avec la plupart des variations utilisées ici ; et même avec les combinaisons qui fonctionnent, la diagonale et le reste de la table ne peuvent pas varier tous les deux, ce qui est précisément nécessaire dans notre cas. Les quelques tests que nous avons pu réaliser montrent cependant que ces modèles pourraient être à la fois d'une grande qualité statistique et d'un grand intérêt sociologique, permettant de développer l'aspect dynamique de notre étude statique réalisée plus bas. Ce problème informatique devra être étudié avec l'aide des auteurs du programme.

23 Comme il est d'usage lorsque l'on utilise des modèles log-linéaires et log-multiplicatifs, nous recourons principalement à l'indicateur BIC (critère d'information bayésien) pour juger de la qualité comparée des modèles. Cet indicateur nous permet de savoir quel est le modèle qui décrit le mieux les données observées en tenant compte de sa parcimonie (nombre de paramètres utilisés) ; mais aussi, par conséquent, de tester de manière statistiquement plus rigoureuse que les indicateurs standard si certains phénomènes observés sont vraiment significatifs au regard des effectifs importants de notre échantillon (Raftery, 1995).

décrivent le mieux l'évolution de l'homogamie sur notre période. Aucune transformation radicale n'a donc eu lieu dans la structure de l'association entre conjoints, mais nous allons voir que l'ampleur du changement net des évolutions structurelles est loin d'être négligeable.

Le double mouvement de l'homogamie d'éducation

L'association quasi-symétrique, qui postule que le diplôme joue le même rôle qu'il soit celui de l'homme ou celui de la femme, est la meilleure manière de décrire l'homogamie d'éducation : ce type d'association est toujours le meilleur en termes de BIC, à part dans un cas (avec une variation complètement libre). Cela signifie que l'hypergamie n'est pas significative sur la période étudiée : on n'observe pas de tendance des hommes à choisir une femme moins éduquée qu'eux, une fois tenu compte du fait que les premiers sont en moyenne plus éduqués que les secondes. Ceci est d'autant plus net qu'inclure dans le modèle un coefficient global mesurant la tendance à l'hypergamie ne l'améliore jamais selon nos tests (non repris ici)[24]. L'éducation semble donc une caractéristique dont la valeur est indépendante du sexe de l'individu : c'est donc un facteur d'homogamie au sens strict, puisqu'il y a symétrie entre les conjoints, plutôt que tendance à aller chercher un conjoint dans un niveau d'éducation adjacent, comme c'est souvent le cas lorsque l'hypergamie des femmes est la règle.

La variation qui décrit le mieux les données est tout aussi intéressante. En fait, deux types de variation sont à peu près équivalents en termes de BIC, et tous deux ont en commun de postuler une variation uniforme tout en permettant à la diagonale et au reste de la table de varier de manière séparée. Le point d'indécision tient donc à la forme de cette variation : linéaire ou non. Puisqu'il est impossible de trancher en l'état des données, et en attendant de voir comment l'homogamie évoluera dans les années à venir, nous représentons les variations non linéaires, sachant qu'il est immédiat pour le lecteur de se représenter la droite qui correspond le mieux aux points de chaque courbe (Figure 6).

Association	Variation								
	Aucune	Libre		Uniforme			Uniforme linéaire		
		E	D	E	D	D, ND	E	D	D, ND
Complète	-1999	0[1]	-2480	-2481[2]	-2615	-2721	-2501[2]	-2632	-2718
Quasi-symétrie	-2176	-1692	-2707	-2695	-2840	**-2955**	-2716	-2856	**-2951**
Lignes-colonnes homogènes[3]	-1836	-[4]	-2305	-2271	-2399	-2378	-2297	-2418	-2421
Distance	-784	-1095	-1243	-1294	-1113	-1271	-1322	-1136	-755
Franchissement	-1267	-1753	-1784	-1848	-1819	-2074	-1873	-1838	-2074

Tableau 14 : Qualité des différents modèles de l'homogamie d'éducation (classification détaillée)

Le modèle d'absence d'association entre diplôme des conjoints présente un BIC, très élevé, de 50 661.

24 Cela ne signifie pas, en revanche, que l'hypergamie n'a jamais existé sur la période ; mais simplement qu'*en moyenne*, elle n'est pas assez forte pour justifier l'ajout d'un coefficient séparé. Nous verrons plus bas (tableau 22, page 69) que l'hypergamie existait pour les cohortes 1969 et 1979, et a disparu après.

Tous les modèles sans exception incluent une série de paramètres pour la diagonale, qui varie soit avec le reste de la table (variante E), soit séparément (variantes D et D, ND).

[1] *Modèle saturé.*

[2] *Ces deux modèles correspondent au modèle UNIDIFF ou à effet de couche, le premier dans sa version de base, le second dans sa version linéaire.*

[3] *Le modèle à lignes-colonnes non homogènes est toujours moins bon.*

[4] *gnm ne prend pas en charge les variations arbitraires dans les scores homogènes ; une tentative réalisée sur des scores non homogènes a cependant montré que ces modèles, même s'ils sont bons, ne sont pas en mesure de rivaliser avec les meilleurs modèles (BIC de -1080).*

Précisons d'emblée que cette variation, *a priori* assez étonnante, est robuste, puisqu'on l'observe en recourant aux trois différentes classifications, qui diffèrent pourtant profondément quant à la définition de l'homogamie stricte qu'elles retiennent (plus le niveau d'agrégation est important, plus la diagonale de la table contient proportionnellement d'individus). D'autre part, ce résultat vient confirmer ceux obtenus plus haut à l'aide de l'indice de diagonalité nette (Figure 5, p. 56), qui présentaient une évolution étonnante : il semble bien que l'homogamie stricte d'éducation, après avoir rapidement diminué jusqu'en 1979, ait stagné, puis soit légèrement remontée après 1999. Ce renforcement est encore difficile à quantifier, puisque les marges d'erreur se sont accrues de manière parallèle[25]. Néanmoins, il est très clair que les deux mouvements ont fortement divergé : d'un côté, l'homogamie stricte se maintient à un niveau sensiblement équivalent à celui de 1969 ; de l'autre, l'homogamie au sens large (éloignement de la situation de choix aléatoire du conjoint) a régulièrement baissé jusqu'à atteindre les deux tiers de son niveau d'origine. En outre, si l'homogamie stricte semble croître à nouveau, celle au sens large continue sa chute rapide.

25 C'est pourquoi nous ne présentons pas le graphique fondé sur la classification détaillée : les effectifs de la diagonale étant dans ce cas trop faibles, les marges d'erreur sont trop importantes pour que les deux courbes soient significativement distinctes chaque année. Mais le fait que l'indicateur BIC porte sa préférence sur un modèle à variation séparée de la diagonale suffit à prouver qu'il y a bien divergence.

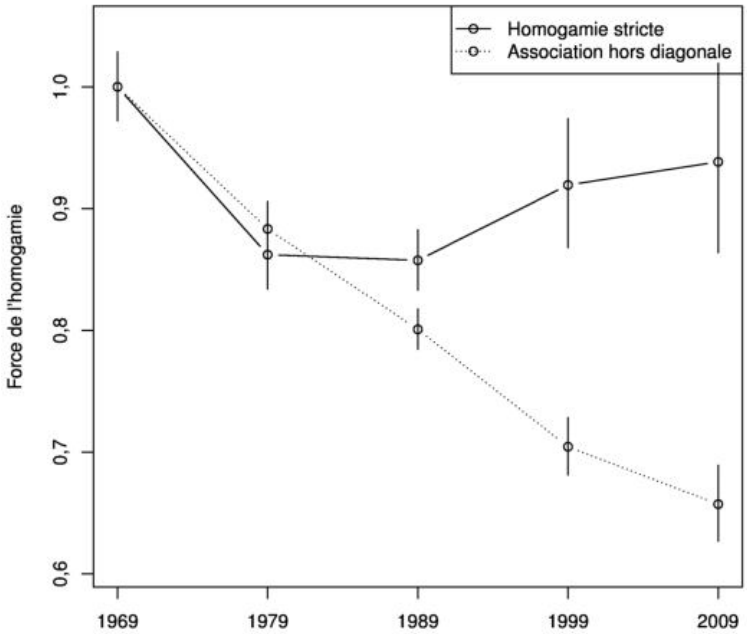

Figure 6 : L'évolution de l'homogamie d'éducation selon le modèle à variation uniforme avec diagonale séparée (classification agrégée)

Champ : couples cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

Cette découverte est d'un grand intérêt : nous n'avons pas connaissance de travaux qui ont eu recours à des modèles à double variation. Vanderschelden (2006a) par exemple, qui fait exception dans la littérature en testant des modèles de constance de l'homogamie et de constance de l'hétérogamie[26], ne combine pas les deux approches, ce qui, avec nos résultats, nous aurait conduits à sélectionner un modèle avec variation de la seule diagonale (Tableau 14). Il est donc possible qu'une partie des divergences constatées entre travaux et entre sociétés soit due à ce double effet : dans une société où la variation la plus forte a été observée sur la diagonale, le meilleur modèle sera celui postulant une variation de celle-ci, alors que ce sera le contraire dans une autre société ; l'évolution observée changera alors de sens. Selon les travaux aussi, un niveau d'agrégation plus ou moins important peut donner la priorité à l'un ou à l'autre des effets. Il serait donc du plus grand intérêt d'appliquer ce type de modèle à des données décrivant différents pays, afin de vérifier si les divergences observées sont bien réelles. Notamment, l'augmentation de l'homogamie d'éducation aux États-Unis n'est peut-être pas aussi opposée au mouvement observé en France, si l'on tient compte du caractère dual de ces deux évolutions.

Comment interpréter l'évolution paradoxale que nous observons ? Elle signifie que d'un côté, les individus ont toujours autant tendance à choisir un conjoint de diplôme identique, mais que, hors de la diagonale de la table, ils vont moins souvent favoriser un conjoint de

26 Que nous préférons désigner par l'expression « homogamie au sens large », puisque l'hétérogamie, qui n'est que le contraire logique de l'homogamie stricte, n'a que peu de sens dans le cadre de modèles mesurant les distances sociales : on peut être plus ou moins hétérogame, selon que l'on choisit un conjoint dans une catégorie plus ou moins éloignée de la sienne, ce qui rend plus correct le recours à une échelle d'homogamie.

diplôme proche, au profit d'individus dont le diplôme est très éloigné du leur. En effet, le tableau 15 nous présente les coefficients correspondant au modèle de quasi-symétrie avec variation uniforme : les coefficients positifs, qui correspondent à des couples plus fréquents qu'attendu, sont clairement placés autour de la diagonale, et les coefficients les plus négatifs, dans les angles de la table. C'est l'écart entre ces coefficients hors-diagonale qui se réduit : tous se rapprochent de zéro (les *odds ratios* qui en découlent tendent, eux, vers l'unité, correspondant à l'indépendance parfaite).

On peut avancer une hypothèse pour expliquer sociologiquement cette divergence. Elle revient en fait à combiner deux théories communément avancées pour prédire le sens d'évolution de l'homogamie : celle de l'obtention d'un statut *(status attainment*), et celle de l'ouverture générale (*general openness*) des sociétés développées, déjà présentées dans la première partie (p. 24). Au lieu de conduire à une croissance puis à une décroissance de l'homogamie d'éducation, la combinaison de ces deux tendances pourrait se traduire par une plus forte ouverture générale envers des individus de diplôme éloigné, mouvement contrecarré sur la diagonale de la table par un renforcement du rôle du diplôme et du temps passé dans le système éducatif. À ce stade, rien ne permet cependant de vérifier cette hypothèse. Néanmoins, un indice en sa faveur est, peut-être, que le statut socio-professionnel, qui n'a pas vu son rôle prendre récemment de l'importance comme l'a fait le diplôme, connaît un affaiblissement général de l'homogamie qui lui correspond – ce que l'on peut voir comme le signe d'un mouvement général d'ouverture auquel l'homogamie stricte d'éducation n'échappe que suite à un développement majeur des enjeux associés au diplôme, qui s'est accompagné d'un accroissement parallèle de le part de diplômés du secondaire et du supérieur.

	Aucun	**CEP**	**BEPC**	**CAP/BEP**	**CAP/BEP+BEPC**	**Bac pro/tech/BP/BT**	**Bac général**	**1er cycle**	**2e et 3e cycle**	**Grande école**
Aucun	**1,22**	**0,42**	**0,29**	-0,81	-0,60	-1,37	-1,23	-1,63	-2,24	-2,71
CEP		**1,19**	**0,64**	-0,75	-0,06	-1,10	-1,07	-1,59	-2,71	-2,89
BEPC			**1,26**	**0,34**	**0,27**	-0,11	-0,37	-0,38	-1,34	-1,65
CAP/BEP				**0,54**	**0,05**	-0,17	-0,38	-0,19	-1,06	-1,51
CAP/BEP+BEPC					**0,64**	-0,42	-0,04	-0,39	-0,88	-1,29
Bac pro/tech/BP/BT						**0,15**	-0,71	-0,24	-0,88	-1,53
Bac général							**0,53**	-0,01	-0,10	-0,51
1er cycle								**0,90**	**0,47**	0,00
2e et 3e cycle									**1,54**	**0,46**
Grande école										**1,28**

Tableau 15 : Coefficients du modèle quasi-symétrique de l'homogamie d'éducation (classification détaillée)

Ce modèle quasi-symétrique mobilise la même association que le modèle à variation uniforme double, mais sans variation temporelle. Des problèmes techniques avec les logiciels gnm et LEM nous ayant empêché de les calculer pour ce dernier modèle, les coefficients présentés ici ont valeur d'illustration de la structure de l'association observée.

Un coefficient positif (respectivement, négatif) signale une fréquence du type de couple correspondant supérieure (inférieure) à ce qui serait attendu en situation de choix aléatoire du conjoint. Dans ce modèle, les positions des hommes et des femmes sont interchangeables.

Tous les coefficients sont significativement différents de zéro au seuil de 1 % exceptés les coefficients de la première ligne (aucune diplôme), ainsi que les paires CEP-CAP/BEP+BEPC, CAP/BEP-CAP/BEP+BEPC, CAP/BEP+BEPC-Bac général, Bac général-1er cycle et 1er cycle-Grande école.

On reproduit ci-dessous les coefficients correspondant à la classification en quatre groupes, qui offrent un aperçu beaucoup plus synthétique. Notons que quelle que soit la classification utilisée, l'association entre conjoints du point de vue de l'éducation se caractérise par une homogamie plus forte aux extrêmes qu'au centre de l'échelle des diplômes. D'autre part, les coefficients hors diagonale sont presque exclusivement négatifs, à l'exception de quelques catégories adjacentes. On verra ci-dessous que l'homogamie socio-professionnelle se structure de manière moins régulière, ce qui correspond au fait que le statut socio-professionnel ne constitue pas une échelle aussi uniforme que celle des diplômes.

	Aucun	**CEP**	**Secondaire**	**Supérieur**
Aucun	**0,95**	**0,04**	-0,30	-2,03
CEP		**0,70**	-0,08	-2,25
Secondaire			**0,73**	-0,19
Supérieur				**0,92**

Tableau 16 : Coefficients du modèle quasi-symétrique de l'homogamie d'éducation (classification en quatre groupes)

Cf. *tableau 15 pour les détails.*

Tous les coefficients sont significativement différents de zéro au seuil de 0,1 %.

L'homogamie socio-professionnelle : un affaiblissement sans ambiguïté

Nous avons appliqué la même procédure aux deux manières auxquelles nous recourons pour mesurer l'homogamie socio-professionnelle : la classe Casmin et la PCS. Dans ce cas, la variation séparée de la diagonale qui décrivait le mieux l'évolution de l'homogamie d'éducation n'est plus nécessaire : une variation uniforme d'ensemble est de meilleure qualité au sens du BIC, signe qu'aucune divergence significative n'existe. La solidité de nos résultats est par ailleurs assurée par le fait que les deux classifications, très différentes (on rappelle que l'on n'utilise ici la classe Casmin qu'en se restreignant aux actifs), donnent lieu aux mêmes conclusions : la baisse observée a été uniforme. Ce mouvement n'est pas très éloigné d'une réduction linéaire au fil des enquêtes, puisque les modèles à variation uniforme et ceux à variation uniforme linéaire sont très proches en termes de BIC. Notons aussi que, contrairement à l'éducation, la catégorie socio-professionnelle reste marquée par des différences de sexe : l'association quasi-symétrique n'a pas une aussi bonne qualité descriptive que l'association complète.

Association	Variation								
	Aucune	Libre		Uniforme			Uniforme linéaire		
		E	D	E	D	D, ND	E	D	D, ND
Complète	**-3214**	0[1]	**-3470**	**-3490**[2]	**-3313**	**-3433**	**-3502**[2]	**-3335**	**-3461**
Quasi-symétrie	**-3092**	**-1588**	**-2952**	**-3375**	**-3202**	**-3307**	**-3388**	**-3224**	**-3341**
Lignes-colonnes homogènes[3]	**-2791**	-[4]	1550	**-3226**	1306	**-3333**	**-3252**	1291	**-3375**
Distance	1675	2019	1798	1516	1574	1565	1503	1557	1524
Franchissement	2667	2707	2757	2484	2565	2512	2472	2544	2480

Tableau 17 : Qualité des différents modèles de l'homogamie de classe Casmin

Le modèle d'absence d'association entre classe Casmin des conjoints présente un BIC, très élevé, de 50 459.

Tous les modèles sans exception incluent une série de paramètres pour la diagonale, qui varie soit avec le reste de la table (variante E), soit séparément (variantes D et D, ND).

[1, 2, 3, 4] Cf. *tableau 14, p. 59.*

Association	Variation								
	Aucune	Libre		Uniforme			Uniforme linéaire		
		E	D	E	D	D, ND	E	D	D, ND
Complète	**-635**	0[1]	**-553**	**-689**[2]	**-654**	**-637**	**-687**[2]	**-652**	**-652**
Quasi-symétrie	**-365**	**-79**	4284	**-426**	**-394**	**-379**	**-422**	4012	**-392**
Lignes-colonnes homogènes[3]	**-250**	-[4]	**-178**	**-633**	**-264**	**-283**	**-630**	**-269**	**-300**
Distance	**-217**	**-42**	**-141**	**-273**	**-263**	**-236**	**-270**	**-257**	**-251**
Franchissement	428	538	496	369	410	410	371	403	389

Tableau 18 : Qualité des différents modèles de l'homogamie de PCS

Le modèle d'absence d'association entre PCS des conjoints présente un BIC, très élevé, de 41 133.

Tous les modèles sans exception incluent une série de paramètres pour la diagonale, qui varie soit avec le reste de la table (variante E), soit séparément (variantes D et D, ND).

[1, 2, 3, 4] Cf. *tableau 14, p. 59.*

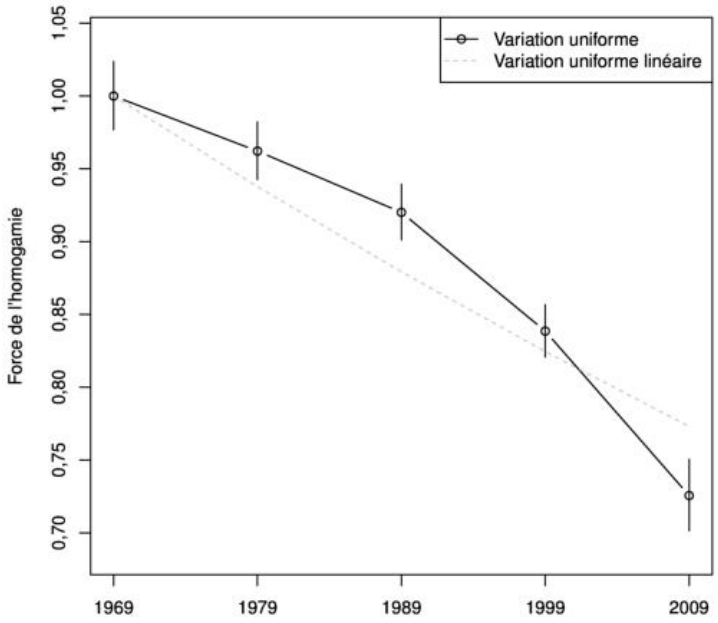

Figure 7 : Évolution de l'homogamie de classe Casmin

Champ : couples d'actifs cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

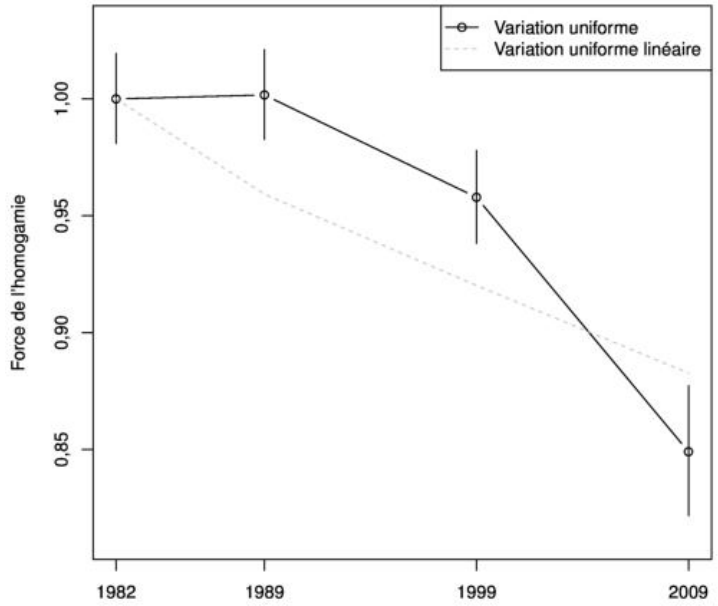

Figure 8 : Évolution de l'homogamie de PCS

Champ : couples cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

Dans les deux cas (figures 7 et 8), le rythme d'affaiblissement de l'homogamie socio-professionnelle semble s'accélérer, même si pour l'instant la variation linéaire décrit aussi bien les données d'après le BIC qu'une variation spécifique à chaque enquête. L'homogamie selon la classe Casmin s'est donc réduite, comme celle d'éducation hors diagonale, d'un tiers, alors que celle de PCS s'est réduite de 15 % (mais sur une durée moindre).

Si l'on s'intéresse aux coefficients d'association des modèles (tableaux 19 et 20), on peut remarquer qu'un certain nombre des coefficients hors diagonale sont positifs, signe d'une proximité entre certains groupes, phénomène que l'on observe beaucoup moins dans le cas de l'homogamie d'éducation (à l'exemple des employés et des ouvriers dans la classification des PCS, ou des employés de niveau inférieur et des ouvriers qualifiés dans la classification Casmin). Nous étudierons plus précisément la structuration de cet espace dans la suite.

Homme	**Femme**										
	I	**II**	**IIIa**	**IIIb**	**IVa**	**IVb**	**IVc**	**V**	**VI**	**VIIa**	**VIIb**
I	**2,91**	**1,29**	**0,38**	-0,56	**0,48**	**0,01**	-0,67	**0,13**	-1,14	-1,49	-1,34
II	**1,33**	**1,44**	**0,37**	-0,07	-0,35	-0,23	-0,65	**0,40**	-0,27	-0,79	-1,18
IIIa	**0,39**	**0,58**	**0,63**	**0,20**	-0,25	-0,30	-0,98	**0,44**	**0,23**	-0,11	-0,83
IIIb	-0,78	-0,14	**0,19**	**0,98**	-0,64	-0,48	-1,16	**0,14**	**0,74**	**0,70**	**0,46**
IVa	**0,41**	-0,34	-0,22	**0,18**	**1,96**	**0,80**	-0,13	-0,38	-0,88	-0,77	-0,63
IVb	**0,09**	-0,37	-0,60	0,00	**2,62**	**2,44**	-0,53	-0,79	-0,75	-0,66	-1,43
IVc	-1,25	-1,40	-1,08	-1,17	**0,04**	-0,40	**5,25**	-1,55	-0,66	**0,32**	**1,90**
V	**1,12**	**0,63**	**0,45**	-0,02	-0,52	-0,39	-0,78	**1,25**	**0,29**	**0,01**	-2,04
VI	-0,77	-0,11	**0,28**	**0,29**	-0,71	-0,28	-1,03	**0,48**	**0,98**	**0,70**	**0,18**
VIIa	-1,45	-0,58	**0,23**	**0,48**	-0,95	-0,55	-0,94	**0,41**	**1,13**	**1,45**	**0,77**
VIIb	-2,00	-1,00	-0,63	-0,32	-1,67	-0,62	**1,62**	-0,52	**0,33**	**0,65**	**4,15**

Tableau 19 : Coefficients du modèle à variation uniforme de l'homogamie de classe Casmin

Un coefficient positif (respectivement, négatif) signale une fréquence du type de couple correspondant supérieure (inférieure) à ce qui serait attendu en situation de choix aléatoire du conjoint.

Le logiciel utilisé (LEM) ne fournit pas d'indication concernant les intervalles de confiance des coefficients des modèles log-multiplicatifs.

	Agricultrice	Indépendante	Cadre	Intermédiaire	Employée	Ouvrière
Agriculteur	**4,35**	-0,82	-1,20	-0,92	-1,10	-0,31
Indépendant	-0,50	**2,12**	-0,09	-0,56	-0,04	-0,93
Cadre	-1,63	-0,04	**2,31**	**0,91**	-0,20	-1,35
Intermédiaire	-1,25	-0,37	**0,51**	**0,70**	**0,27**	**0,14**
Employé	-0,89	-0,36	-0,24	**0,24**	**0,75**	**0,51**
Ouvrier	-0,08	-0,53	-1,29	-0,36	**0,33**	**1,93**

Tableau 20 : Coefficients du modèle à variation uniforme de l'homogamie de PCS

Cf. *tableau 19 pour les détails.*

Du point de vue de l'homogamie socio-professionnelle, nos résultats sont donc contradictoires avec ceux de Vanderschelden (2006a, p. 45), qui concluait à une stabilité globale de l'homogamie de PCS, seules certaines associations variant, dans un sens comme dans l'autre. En revanche, ils rejoignent ceux de Vallet (1986, pp. 687–688), qui observait une diminution de l'homogamie socio-professionnelle entre 1962 et 1982 (à partir de sondages au vingtième du recensement de la population), de l'ordre d'un sixième de sa réduction brute.

En ce qui concerne l'homogamie d'éducation, nos conclusions sont relativement compatibles avec celles de Vanderschelden (2006a, pp. 49–53), puisque son article concluait que l'homogamie nette d'éducation avait bien décru, mais de manière irrégulière : certains niveaux et certaines périodes avaient vu l'homogamie augmenter ; cette hypothèse, testée ici à l'aide du modèle à forme de régression, n'est cependant pas confirmée par nos données. Confirmant ce qu'on a dit plus haut, nos résultats s'opposent surtout à ceux de Forsé et Chauvel (1995, p. 137), qui observaient une stabilité de l'homogamie d'éducation entre les deux générations étudiées.

Afin d'expliquer ces divergences, il est important de noter que l'étude de Forsé et Chauvel et celle de Vanderschelden se fondent chacune sur une seule enquête, la variable temporelle étant l'âge, alors que celle de Vallet recourait, comme la nôtre, à plusieurs enquêtes successives. Étant donnée la faiblesse, en termes d'effectifs, des évolutions observées, il n'est pas impossible que le biais lié à l'attrition des couples suffise à expliquer cette divergence de résultats – ce dont Vanderschelden ne se cache pas dans les pages citées. En revanche, la différence entre couples mariés et couples cohabitants ne semble pas à l'origine des différences de conclusions : les travaux de Vallet et de Forsé et Chauvel ne n'étudiaient que les couples mariés, alors que Vanderschelden, retenait comme nous toutes les unions ; le jeu des oppositions et des convergences échappe donc aux camps que pourrait ébaucher ce choix méthodologique.

Modèle	Diplôme (détaillé)				Classe Casmin actifs				PCS			
	D.L.	L^2	Δ%	BIC	D.L.	L^2	Δ%	BIC	D.L.	L^2	Δ%	BIC
Absence d'association	405	55 391	25,4	50 661	500	56 062	31,2	50 459	100	42 271	23,4	41 133
Association stable	360	2 029	4,3	-2 176	400	1 268	3,7	-3 214	75	219	1,3	-635
Variation uniforme	356	1 463	3,4	-2 695	396	947	3,2	-3 490	72	131	0,9	**-689**
Variation uniforme linéaire	359	1 478	3,5	-2 716	399	969	3,3	**-3 502**	74	156	1,0	-687
Variation uniforme double[1]	343	1 051	2,9	**-2 955**	382	848	2,9	-3 433	68	91	0,7	-637

Tableau 21 : Détails des modèles de référence pour les trois types d'homogamie

L'association entre caractéristiques des conjoints est quasi-symétrique dans le cas du diplôme, complète dans les cas de la classe Casmin et des PCS.

On a mis en gras la valeur du BIC qui est la plus négative pour chaque dimension.

D.L. : degrés de liberté. L^2 : déviance du modèle par rapport au modèle saturé (-2 × log-vraisemblance). Δ% : indice de dissimilarité entre les prédictions du modèle et les données observées. BIC : critère d'information Bayésien (-2 × log-vraisemblance - D.L. × log(N_{obs}).

[1] Variation séparée de la diagonale et du reste de la table.

IV Les transformations de l'espace social

Les modèles présentés jusqu'ici nous ont permis de mesurer la transformation de la structure et du niveau de l'association entre caractéristiques des conjoints dans le temps. Il n'est cependant pas inintéressant d'étudier l'espace social tel que le dessine l'homogamie de manière statique, c'est-à-dire séparément à chaque enquête. En effet, les modèles que nous avons utilisés pour décrire les variations de l'homogamie ne permettaient pas à l'espace social de changer profondément de structure : même si la qualité de leurs prédictions garantit qu'ils mesurent bien une évolution réelle en termes de niveau, il est possible que des modèles plus parcimonieux donnent, à chaque enquête, des interprétations plus riches.

Les modèles qui sont apparus comme les meilleurs jusqu'ici mettaient en jeu des descriptions assez détaillées de l'association entre conjoints : association complète dans le cas de l'homogamie socio-professionnelle, association quasi-symétrique dans celui de l'homogamie d'éducation. Au contraire des spécifications de type distance, franchissement, ou association lignes-colonnes, ces associations ne faisaient pas d'hypothèse spécifique concernant la structuration de l'espace social. Grâce à leur précision, elles sont plus à même de décrire un espace social qui se transformerait, alors que des modèles plus parcimonieux, mais reposant sur des hypothèses plus fortes auraient du mal à s'adapter aux enquêtes successives, puisqu'ils décriraient bien certaines d'entre elles, correspondant à un état donné de la société, mais très mal les autres. Comme les modèles dynamiques ne permettaient pas

d'utiliser un type d'association pour certaines enquêtes, et un autre plus tard, les associations les plus riches étaient mécaniquement désavantagées[27]. Nous allons donc étudier ici les transformations, non pas du niveau de l'homogamie, mais de la forme de l'association, ce qui ne remet pas en cause les mesures effectuées plus haut, mais les enrichit.

La démarche est simple : au lieu de combiner chaque type d'association avec une manière différente de la faire varier dans le temps, on applique séparément un modèle statique d'association à chaque enquête. On recourt à tous les types d'association présentés plus haut : complète ; quasi-symétrique ; distance ; franchissement ; association lignes-colonnes à une, deux, trois ou quatre dimensions. Pour chaque type (excepté le premier, déjà saturé), on teste aussi l'intérêt d'ajouter un coefficient d'hypergamie, mesurant l'asymétrie hommes-femmes. Pour chaque manière de mesurer l'homogamie, on présente la valeur du BIC obtenue pour tous les modèles qui se placent au moins une fois parmi les trois premiers (tableaux 22, 23 et 24).

L'éducation : nouvelles oppositions, fin de l'hypergamie

	Lignes-colonnes homogènes	Lignes-colonnes homogènes avec hypergamie	Lignes-colonnes homogènes, deux dimensions	Lignes-colonnes homogènes, deux dimensions avec hypergamie	Lignes-colonnes homogènes, trois dimensions
1969	-317	-315	-409	**-416**	-383
1979	-387	-388	-412	**-418**	-363
1989	-430	-420	**-445**	-435	-426
1999	-352	-342	**-386**	-376	-357
2009	-307	-299	**-342**	-333	-326

Tableau 22 : Les trois meilleurs modèles statiques de l'homogamie d'éducation

Ces modèles sont ceux qui obtiennent l'un des trois meilleurs scores BIC au moins une année.

Les modèles de l'homogamie d'éducation qui apparaissent comme les meilleurs à chaque enquête sont bien ceux que l'on soupçonnait : les différentes variantes d'association lignes-colonnes. Le tableau 22 nous montre ainsi que l'espace matrimonial des diplômes est le mieux décrit par deux dimensions homogènes : cela signifie, en cohérence avec le fait que le modèle de quasi-symétrie avait été retenu plus haut, que les diplômes ont la même valeur pour les hommes et pour les femmes, mais qu'ils ne se laissent pas classer, comme on est souvent enclin à le penser, sur une échelle unique.

27 À ce problème qui est intrinsèque à l'utilisation de modèles dynamiques, s'ajoute la limitation technique qui nous a empêché de faire varier dans le temps les scores des modèles d'association lignes-colonnes, ce qui aurait limité l'ampleur du problème. Cette possibilité n'aurait cependant pas permis de découvrir, ainsi que nous le verrons dans cette partie, que le nombre de dimensions structurant l'espace social a varié.

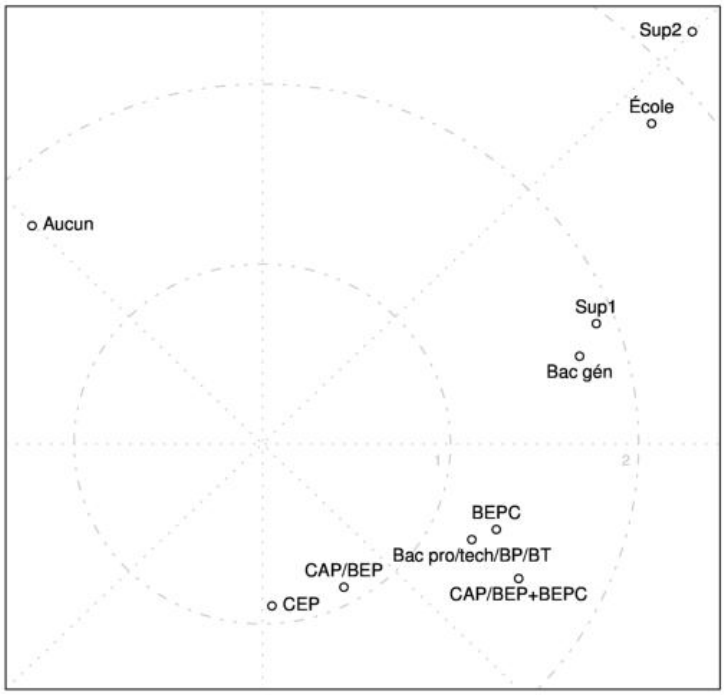

Figure 9 : L'espace matrimonial des diplômes en 1969

Champ : couples cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

Malgré ce trait commun – l'existence d'une double échelle –, l'espace des diplômes s'est profondément transformé en quarante ans. Outre la disparition de l'hypergamie qu'atteste le fait qu'à partir de 1989, l'ajout d'un tel paramètre n'améliore plus les modèles (point sur lequel nous reviendrons), la structuration de l'espace a elle-même profondément changé. En 1969 (figure 9, *cf.* aussi l'encadré « La représentation des modèles lignes-colonnes », p. 72), la droite de la figure développe l'échelle des diplômes, l'augmentation du niveau allant de pair avec un plus grand degré de fermeture (éloignement progressif de l'origine). Les non diplômés s'opposent au reste des catégories, du fait de leur situation très isolée (aucun hémisphère commun avec une autre catégorie). En 2009, en revanche (figure 10), la configuration est bien différente : les non diplômés on rejoint le groupe des diplômes inférieurs. Deux échelles différentes relient le bas et le haut de la hiérarchie des diplômes : par le nord-ouest, l'enseignement professionnel ; par le sud-est, l'enseignement général. On observe ainsi une double hiérarchie des diplômes, correspondant à deux univers institutionnels, culturels et professionnels. À mesure que l'on s'approche des extrémités cette hiérarchie, on s'éloigne aussi du centre de la figure, signe d'une plus grande fermeture des groupes concernés.

Entre 1969 et 2009, donc, les non diplômés ont progressivement réintégré l'échelle de valeur des diplômes, en se plaçant au niveau des titulaires du certificat d'études ; parallèlement à ce mouvement, l'échelle des diplômes s'est scindée selon la distinction entre professionnel et général. Nous pouvons donc avancer que nous observons ici les effets du mécanisme déjà décrit lors de la discussion du rôle des catégories utilisées : à mesure que les couches sociales jusqu'alors non diplômées ont bénéficié de la massification scolaire, et notamment du développement de l'enseignement professionnel, la catégorie des sans diplôme,

qui représentait une grande partie de la population, s'est réduite, entraînant une transformation de la position des catégories qui ont accueilli les populations concernées, et finalement de l'espace global des diplômes.

L'apparition d'une opposition entre enseignement général et enseignement professionnel n'est cependant pas uniquement le fait des diplômes inférieurs : les titulaires d'un diplôme du premier cycle de l'enseignement supérieur apparaissent du côté de l'enseignement professionnel, et ce malgré une position élevée sur le premier axe. La position de ce groupe est due au fait que ses membres ont plus tendance que, par exemple, les titulaires d'un baccalauréat général ou d'un diplôme de deuxième ou troisième cycle, à choisir un conjoint parmi les titulaires d'un diplôme professionnel ; ce qui n'est pas contradictoire avec le fait qu'il entretiennent par ailleurs une grande proximité avec les deux premiers groupes cités. Affiner cette analyse exigerait de diviser les diplômes du premier cycle universitaire en sous-groupes, en distinguant notamment les filières universitaires dites « courtes » (BTS et DUT) des diplômes généraux ; mais aussi de se prémunir contre des effets de génération, la valeur sociale d'un diplôme de premier cycle ayant pu connaître d'importantes transformations entre les quinquagénaires et les trentenaires que notre table d'homogamie réunit sous une même étiquette[28].

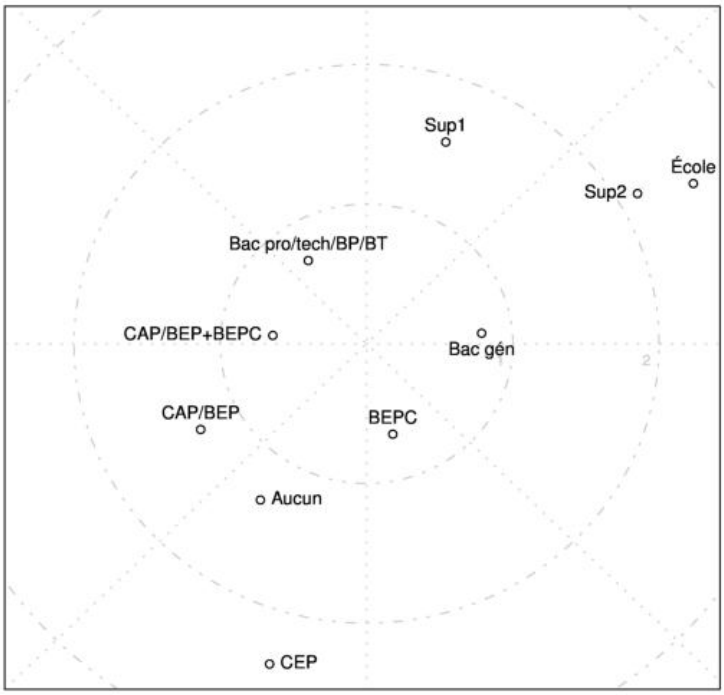

Figure 10 : L'espace matrimonial des diplômes en 2009

Champ : couples cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

Indépendamment de la question du premier cycle du supérieur, cette structure de l'espace social des diplômes est d'une grande importance théorique, pour la description du système scolaire autant que pour la compréhension des mécanismes du choix du conjoint. En effet,

28 Au même titre que, du côté de l'enseignement général, la catégorie certificat d'études (CEP), qui, ne comprenant pour les enquêtes récentes que des individus appartenant aux générations les plus anciennes, apparaît particulièrement liée aux non diplômés du fait, en partie du moins, d'un artefact.

l'existence d'une seconde structuration, clairement distincte de celle constituée par la valeur du diplôme – indicateur selon la théorie économique du marché matrimonial d'un statut social et d'un niveau de vie futurs (Becker, 1991) –, montre que le choix d'un conjoint met en jeu des processus plus complexes qu'un simple échange. Ces premiers résultats tendraient donc à venir en renfort de la théorie de la similarité, qui explique le choix du conjoint par une convergence de modes de vie et de pensée entre individus (Blau, 1967). Mais des stratégies de distinction et de reproduction (Bourdieu, 1972) pourraient tout aussi bien expliquer ce phénomène : si l'on suppose que l'enseignement général bénéficie d'un prestige supérieur lié à un capital culturel plus élevé, on peut comprendre qu'une seconde dimension du choix du conjoint apparaisse, en plus de la structuration selon le niveau de capital global dont un titre scolaire est à la fois le signe et la condition.

La représentation des modèles lignes-colonnes

Les représentations en deux dimensions des espaces matrimoniaux présentées ici, si elles ont l'avantage de synthétiser les proximités entre groupes, ne sont néanmoins pas d'une interprétation immédiate. Il est notamment essentiel d'éviter l'erreur qui consiste à considérer la distance euclidienne (la plus naturelle) entre les points correspondant à deux catégories comme la distance sociale entre ces catégories, telle que spécifiée par le modèle d'association lignes-colonnes : le lien entre ces deux notions est plus complexe. Ce type de représentation est en fait apparenté aux graphiques obtenus par analyse des correspondances (Goodman, 1986), la principale différence étant qu'ici, toutes les dimensions de l'analyse des correspondances sont synthétisées en un ou deux axes seulement, ce qui interdit de leur donner un sens univoque.

Dans sa version à une dimension, ce type de modèle prédit les effectifs d'un type de couple donné par la formule $\ln n_{ij} = n_i^H + n_j^F + \delta_i \delta_j$, n_{ij} correspondant à la part des couples dans lesquels l'homme est de la catégorie i et la femme de la catégorie j, n_i^H et n_j^F à la proportion de chacune de ces catégories chez les hommes et les femmes, δ_i et δ_j aux scores calculés pour les catégories en question (identiques, à catégorie donnée, pour les hommes et pour les femmes)[29]. Ainsi, en l'absence d'association entre caractéristiques des conjoints, le dernier terme de l'équation est nul, ce qui correspond à une fréquence du type de couple considéré égale au produit de la proportion des catégories i et j au sein des populations respectives des hommes et des femmes (situation d'indépendance, ou de choix aléatoire du conjoint). À l'inverse, plus le produit des scores est important en valeur absolue, et plus l'association est de grande ampleur ; le signe du produit détermine alors si cette association correspond à une attraction (sur-représentation) ou à une répulsion (sous-représentation).

On comprend donc quelle est la relation que doivent entretenir les scores de deux catégories entre eux pour qu'elles soient considérées comme socialement proches : il faut que leurs scores soient non seulement de même signe, mais aussi qu'ils soient tous les deux élevés. Dans l'autre sens, deux catégories très éloignées ont toutes deux des scores élevés, mais de signes opposés. Cette représentation est largement contre-intuitive : deux catégories proches sur l'axe seront moins fortement associées si elles se trouvent près de l'origine que deux autres catégories géométriquement plus éloignées, mais qui, éloignées de l'origine, ont des scores très élevés (et de même signe). Pour le dire autrement, des catégories qui sont séparées par la même distance (et du même côté de l'axe) seront d'autant plus fortement associées qu'elles sont éloignées de l'origine,

29 Sur les figures, les scores des deux dimensions sont normalisés de sorte que, sur chaque dimension, et en pondérant les scores par les effectifs de leurs catégories, leur somme soit nulle, et la somme de leurs carrés égale à un : *cf.* Goodman (1979, p. 548, éq. 4.8) ou encore Agresti (2002, p. 380, éq. 9.15). Contrairement à la version proposée par ces auteurs, cependant, le modèle utilisé ici n'inclut pas de paramètre multiplicatif indiquant la force générale de l'association sur chaque axe, ce qui permet de représenter un espace homogène, les coordonnées incluant toute l'information.

condition requise pour que le produit de leurs scores soit maximal. Ce mécanisme fonctionne de manière similaire pour des catégories socialement éloignées : elles doivent être situées sur des côtés opposés de l'axe, et loin de l'origine.

Enfin, notons que la position d'une catégorie donne, au-delà de ces comparaisons deux-à-deux, un indice de son association avec les autres. En effet, un point situé loin de l'origine est le signe d'une forte attraction avec les points situés de son côté de l'axe, et d'une forte répulsion à l'égard de ceux situés du côté opposé. Au contraire, un point situé très près de l'origine signale une catégorie moyenne, qui entretient des relations moins contrastées avec les autres, quel que soit le côté de l'axe sur lequel elle se situe – même si elle peut être très fortement associée à une catégorie qui serait assez extrême pour compenser la faiblesse de son propre score.

Ce schéma se complique encore légèrement lorsque deux dimensions sont en jeu : dans ce cas, l'équation devient $\ln n_{ij} = n_i^H + n_j^F + \delta_i \delta_j + \gamma_i \gamma_j$, ce dernier terme représentant le produit des scores de la seconde dimension. Une association positive sur un axe peut alors être compensée par une autre, négative, sur le second. Pour cette raison, il convient de ne pas aborder la lecture de ces graphiques en se fondant sur les coordonnées des points sur les deux axes, mais de manière circulaire – technique qui rend l'interprétation des figures beaucoup plus naturelle. En effet, l'association entre deux catégories s'écrit $\delta_i \delta_j + \gamma_i \gamma_j = d_i \cos\theta_i . d_j \cos\theta_j + d_i \sin\theta_i . d_j \sin\theta_j$, avec d_i et d_j les distances des points à l'origine, θ_i et θ_j leur angle. En choisissant conventionnellement l'angle du premier point comme l'angle nul, on obtient $\delta_i \delta_j + \gamma_i \gamma_j = d_i d_j \cos\theta_j$. L'association entre deux catégories correspond donc au produit des distances à l'origine de leurs deux points, multiplié par le cosinus de l'angle formé par les segments reliant les deux points à l'origine.

Ce détour mathématique nous permet donc d'énoncer une méthode simple d'interprétation : un point est positivement associé avec tous les points situés dans le même hémisphère de la figure, négativement avec les autres, la frontière correspondant à la perpendiculaire au segment reliant le point à l'origine ; les points situés sur cette frontière ont un coefficient d'association nul avec le groupe étudié, correspondant à une situation aléatoire de choix entre les deux catégories. D'autre part, plus un point est éloigné de l'origine, plus il sera contrasté dans ses associations : deux points situés à des angles proches, et éloignés de l'origine, seront fortement associés ; deux points diamétralement opposés (l'expression courante rejoignant ici la trigonométrie) seront d'autant plus fortement opposés qu'ils seront éloignés de l'origine, chacun dans son sens. Entre ces deux extrêmes, la force de l'association dépendra à la fois de la proximité angulaire, et de l'éloignement de l'origine.

Sociologiquement, cette représentation est en fait assez illustrative. L'angle au centre permet de situer les groupes sur une forme d'échelle circulaire, rapprochant les groupes associés, opposant ceux qui se repoussent. La distance à l'origine permet, elle, de mesurer la fermeture générale des groupes : au centre, les catégories les plus moyennes et ouvertes, à la périphérie les groupes les plus isolés. Moins intuitive qu'une représentation sur deux axes comme l'analyse des correspondances, cette présentation présente donc de réels avantages dans le cas de la sociologie de l'homogamie.

Relevons un autre point d'intérêt concernant l'homogamie d'éducation : alors que jusqu'en 1979, l'inclusion d'un paramètre d'hypergamie améliorait le modèle, ce n'est plus le cas depuis 1989. Nous pouvons donc conclure que même une fois tenu compte de la structure de diplôme des populations respectives des hommes et des femmes – qui a évolué, on l'a dit, jusqu'à voir les femmes devenir plus diplômées que les hommes –, la tendance des femmes à choisir des hommes plus éduqués qu'elles (ou inversement) a complètement disparu dans la société française. Ce phénomène ne s'est pas non plus inversé (le paramètre aurait alors pris une valeur négative), ce qui était une possibilité. Cette évolution peut s'interpréter, de manière positive, comme le signe d'une progression de l'égalité des sexes.

Elle peut aussi correspondre à une plus grande importance du diplôme dans la société, qui rend de plus en plus coûteux de choisir une conjointe de diplôme inférieur plutôt que de même niveau scolaire. Si l'on accepte un cadre d'analyse dans lequel les individus sont en concurrence pour des conjoints de statut élevé (au moins sur la première dimension que nous avons identifiée plus haut), un individu ne peut espérer viser plus haut que son propre niveau, puisque les conjoints potentiels de niveau supérieur sont déjà « réservés » (*modulo* les asymétries hommes-femmes au sein de chaque groupe de diplôme) par les individus du sexe opposé qui sont du même niveau. Si un individu veut maximiser le statut social de son conjoint, il doit donc former un couple homogame. Le fait que ce mécanisme d'optimisation sous contrainte soit symétrique explique que l'on aboutisse, sous ces hypothèses, à une homogamie potentiellement parfaite.

Ce modèle hypothétique peut constituer une explication supplémentaire de la divergence que nous avons observée plus haut (*cf.* page 61) entre homogamie stricte et homogamie large, seule la seconde diminuant depuis quarante ans : si la norme d'hypogamie de l'homme disparaît au profit d'une tendance à préserver son statut social, l'homogamie au sens strict se renforce ; mais cette évolution n'est pas incompatible avec une tendance générale à l'ouverture sociale. Les deux processus peuvent se combiner pour aboutir à une stabilité de l'homogamie stricte, tout en observant plus fréquemment des rencontres « erratiques » entre deux individus de diplômes très différents.

L'unification de l'espace socio-professionnel

Nous pouvons tirer des enseignements similaires concernant l'homogamie socio-professionnelle. L'hypergamie n'a dans ce cas jamais été la règle (du moins, pas au niveau d'agrégation utilisé), ce n'est donc pas cette transformation que nous observons. En revanche, la structuration de l'espace des classes Casmin suit une évolution originale (tableau 23). Alors que cet espace se décrit le mieux à l'aide de trois dimensions jusqu'en 1999, il n'en requiert plus que deux en 2009. Ce changement tient en fait à la quasi-disparition du groupe des ouvriers agricoles (VIIb) : dans l'espace à trois dimensions, le dernier axe opposait toujours de manière très nette ce groupe à celui des indépendants sans salariés (IVb). Du côté des premiers, on trouvait les agriculteurs (IVc) et les ouvriers non qualifiés hors agriculture (VIIa) ; de celui des seconds, les indépendants avec salariés (IVa) et les classes supérieures (I). Cette troisième dimension n'est plus nécessaire à la description de l'espace social en 2009, probablement parce que d'une part les indépendants se sont rapprochés de la moyenne sur cet axe, et d'autre part parce que les ouvriers agricoles sont trop peu nombreux pour justifier un axe à eux seuls.

	Lignes-colonnes, deux dimensions	**Lignes-colonnes homogènes, deux dimensions**	**Lignes-colonnes, trois dimensions**	**Lignes-colonnes homogènes, trois dimensions**
1969	-367	-289	-317	**-478**
1979	-321	-199	-300	**-575**
1989	-410	-436	**-462**	-408
1999	-363	-380	-317	**-681**
2009	-404	**-514**	-313	-475

Tableau 23 : Les trois meilleurs modèles statiques de l'homogamie de classe Casmin

Ces modèles sont ceux qui obtiennent l'un des trois meilleurs scores BIC au moins une année.

Si l'on se tourne vers les deux autres axes, on peut observer que l'espace des classes Casmin s'est considérablement transformé en quarante ans. La figure 11 représente les deux premiers axes de cet espace en 1969 : on distingue clairement l'échelle du salariat, d'en haut à gauche au centre droit, alors que l'autre hémisphère développe l'échelle des indépendants, à laquelle se rattachent les ouvriers agricoles (VIIb), du fait d'une proximité avec les agriculteurs (IVc), et les classes supérieures (I), du fait d'une proximité avec les indépendants avec salariés (IVa). On rappelle que le troisième axe rapproche les classes supérieures (I) et les indépendants avec salariés (IVa), et les éloigne des autres, réordonnant ainsi les groupes de cet hémisphère selon leur statut social. Les couches supérieures des employés (IIIa) et les techniciens, groupes moyens par excellence, forment déjà un groupe central, peu contrasté dans ses associations, bien que légèrement du côté des salariés.

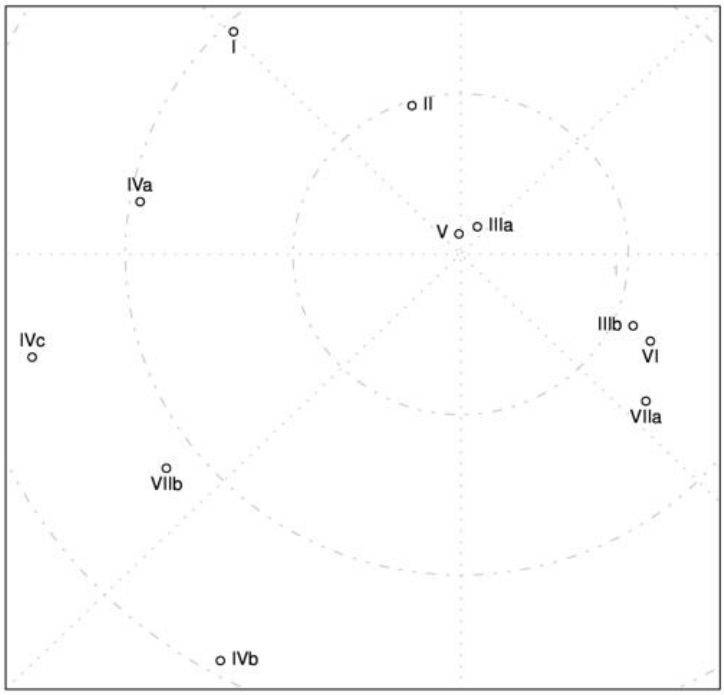

Figure 11 : L'espace matrimonial des classes Casmin en 1969 (deux premiers axes)

Champ : couples d'actifs cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

En 2009, l'espace, à seulement deux dimensions, n'est plus construit autour de l'opposition salariés-indépendants de la même manière (figure 12). Deux « sentiers » d'opposition se dessinent, désormais reliés au centre par un groupe pivot, toujours formé des techniciens (V) et des couches supérieures des employés (IIIa) : de l'angle inférieur gauche à l'angle supérieur droit, l'échelle des indépendants ; sur l'autre diagonale, l'échelle des salariés. Les indépendants se découpent en deux groupes : d'un côté, les indépendants, avec ou sans salariés (IVa et IVb), de l'autre, les agriculteurs (IVc), auxquels se rattache le petit groupe des ouvriers agricoles (VIIb), qui s'associe aussi fortement aux classes populaires Les salariés suivent leur hiérarchie, avec un pôle populaire en haut à gauche, un pôle supérieur en bas à droite. Il n'existe plus de noyau des indépendants : ceux qui n'ont pas de salariés (IVb) se sont rapprochés du groupe central ; ceux qui en ont (IVa) deviennent, avec une position éloignée du

centre, un groupe assez contrasté dans ses associations : fortement opposé aux catégories de la gauche de la figure, il s'associe fortement avec les catégories supérieures (I, et dans une moindre mesure II).

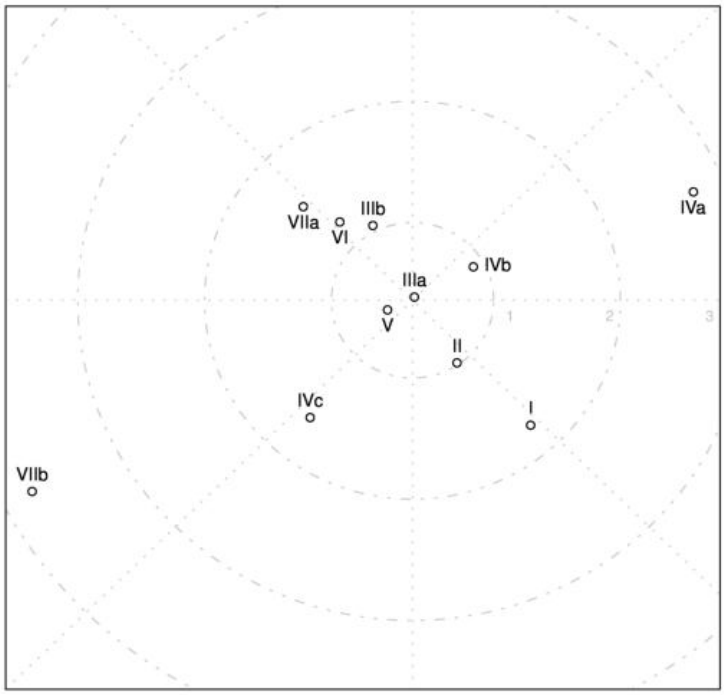

Figure 12 : L'espace matrimonial des classes Casmin en 2009

Champ : couples d'actifs cohabitants dans lesquels l'homme est âgé de 30 à 59 ans à la date de l'enquête.

La situation est plus difficile à interpréter en ce qui concerne les PCS. Du fait du petit nombre de catégories utilisées par cette classification (6, contre 11 pour la classe Casmin[30]), les modèles complexes ne diffèrent en pratique que par quelques coefficients, ce qui favorise l'instabilité du choix du meilleur modèle d'une année sur l'autre (tableau 24). On peut cependant retenir de ces modèles que l'espace des PCS se structurait, au moins jusqu'en 2009, en deux dimensions ; en 2009, le modèle d'association lignes-colonnes homogènes en deux dimensions, s'il n'est pas le meilleur, en reste cependant relativement proche, ce qui invite à rester très prudent. L'hypergamie, si elle a vraiment existé en 1982, a en tout cas disparu depuis.

Néanmoins, si l'on se fie au modèle décrivant le mieux l'homogamie de PCS en 2009, le modèle de distance, l'espace matrimonial des PCS est assez simple : il se structure selon une hiérarchie valable pour les deux sexes (cadres et professions intellectuelles supérieures, professions intermédiaires, indépendants, employés, agriculteurs, ouvriers)[31], avec quatre niveaux de probabilité de mise en couple hétérogame seulement. Ainsi, choisir un conjoint au

30 Il serait bien sûr intéressant de reproduire cette analyse en se fondant sur le niveau 2 des PCS, qui comporte 42 catégories.

31 Cette hiérarchie, qui peut sembler peu naturelle, est celle qui correspond le mieux aux données, étant obtenue à partir des scores d'un modèle lignes-colonnes homogènes à une dimension (*cf.* note 21, p. 57).

sein de l'une des deux catégories adjacentes à son propre groupe dans chaque direction (vers le haut ou vers le bas de la hiérarchie) présente la même probabilité au sens de l'*odds ratio* pour tous les groupes (*cf.* les résultats présentés dans le tableau 25). Au-delà de deux catégories franchies dans la même direction, chaque nouvelle barrière donne lieu à un *odds ratio* distinct, le même pour tous les groupes de départ. Notons que ces probabilités ne sont valables que pour les groupes qui, situés à l'extrémité de la hiérarchie, peuvent effectivement traverser ce nombre de catégories : un employé étant situé au milieu de l'échelle ne peut traverser, dans chaque direction, que trois catégories au maximum.

	Lignes-colonnes					Distance
		Deux dimensions				
	Une dimension	Hétérogènes		Homogènes		
		Sans hypergamie	Avec hypergamie	Sans hypergamie	Avec hypergamie	
1982	-34,9	-54,8	27,5	-86,4	**-90,2**	37,5
1989	-43,6	-44,7	39,8	**-102,1**	21,9	81,1
1999	**-62,4**	-25,5	-25,8	-13,9	-6,9	-11,0
2009	-42,7	3,9	-46,8	-52,5	-44,1	**-64,5**

Tableau 24 : Les trois meilleurs modèles statiques de l'homogamie de PCS

Ces modèles sont ceux qui obtiennent l'un des trois meilleurs scores BIC au moins une année.

Dans les deux cas, classe Casmin et PCS, l'espace socio-professionnel tel qu'observé à travers les choix de mise en couple s'est considérablement simplifié en quarante ans : l'hypergamie a disparu ; le nombre de dimensions s'est réduit, passant dans le premier cas de trois à deux, dans le second de deux à une seule (modèle de distance hiérarchisé). Il semble que cette transformation soit notamment due à l'affaiblissement du clivage entre indépendants et salariés (que l'on observe aussi bien avec la classe Casmin qu'avec la PCS), par le biais de la « moyennisation » du groupe des indépendants sans salariés, et à la réduction des effectifs du pôle agricole (agriculteurs exploitants et ouvrier agricoles).

	C.	Int.	Ind.	E.	A.	O.
Cadre	**0,95**	**0,07**	**0,07**	-0,60	-0,92	-1,78
Intermédiaire	**0,07**	**0,18**	**0,07**	**0,07**	-0,60	-0,92
Indépendant	**0,07**	**0,07**	**2,15**	**0,07**	**0,07**	-0,60
Employé	-0,60	**0,07**	**0,07**	**0,60**	**0,07**	**0,07**
Agriculteur	-0,92	-0,60	**0,07**	**0,07**	**5,14**	**0,07**
Ouvrier	-1,78	-0,92	-0,60	**0,07**	**0,07**	**0,24**

Tableau 25 : Coefficients du modèle de distance de l'homogamie de PCS (2009)

Un coefficient positif signale une sur-représentation du type de couple considéré par rapport à la situation de choix aléatoire du conjoint. Les positions des hommes et des femmes sont interchangeables dans ce modèle.

Tous les coefficients sont significativement différents de 0 au seuil de 0,1 % à part les coefficients d'homogamie Intermédiaire × Intermédiaire, Ouvrier × Ouvrier, et le premier

niveau de distance (0,07), qui peut donc aussi être considéré comme la référence.

Ce modèle, simplifié par rapport au modèle de distance plus générique utilisé à titre de test dans le tableau 24, présente un BIC encore amélioré de -73,5 (déviance de 78 pour 16 degrés de liberté, indice de dissimilarité de 1,99 %).

Analyse biographique de l'homogamie : scolarité et mise en couple

L'étude de l'homogamie s'est traditionnellement fondée sur des mesures transversales, dans un premier temps à l'aide de simples tables carrées croisant les caractéristiques des conjoints, puis en soumettant ces tables à des modèles log-linéaires ou log-multiplicatifs plus complexes, souvent avec une dimension temporelle tenant à une évolution inter-cohortes. Plus récemment cependant, des approches biographiques ont été mises en œuvre, visant à ouvrir la boîte noire de l'homogamie, afin de comprendre les processus qui y mènent plutôt que de se contenter de mesurer son résultat – qui constitue déjà un sujet d'étude en lui-même. Dans l'ouvrage collectif déjà cité, Blossfeld et Timm (Blossfeld et Timm, 2003b) essaient notamment de mesurer quelle est l'influence des systèmes éducatifs sur la mise en couple de manière homogame (ou non), considérant ceux-ci comme des marchés matrimoniaux, et dans l'idée que l'allongement des études et la massification scolaire peuvent mener à un renforcement de l'homogamie de diplôme. Nous essaierons ici de reproduire leur démarche générale, tout en nous en éloignant pour tenter d'ébaucher des critiques à l'encontre de la méthode et des résultats obtenus, dans la contribution concernant la France, par Goux et Maurin (2003).

Nous avons la chance de disposer pour cela de données parfaitement adaptées à notre objectif : l'enquête Biographies et entourage a été menée en 2000-2001 sur un échantillon de 2830 personnes représentatives de la population d'Île de France et nées entre 1930 et 1950. Le questionnaire de cette enquête interroge les individus sur leur parcours personnel et sur celui de leur entourage de manière longitudinale, à l'aide de fiches retraçant la chronologie des événements. Nous connaissons donc en particulier les caractéristiques des conjoints successifs des interrogés *à la date de mise en couple*, ainsi que la situation de l'individu à cette date, notamment le moment de fin d'études. Ceci nous permet donc d'éviter un écueil majeur rencontré par Goux et Maurin dans leur travail fondé sur l'enquête Formation et qualification professionnelle de 1993, qui supposaient, faute de mieux, que le premier conjoint et le conjoint actuel partageaient les mêmes caractéristiques.

Nous commencerons par étudier de manière non paramétrique les rythmes de mise en couple, pour progressivement tenter d'appliquer des modèles plus puissants, mais qui exigent de faire des hypothèses assez fortes, dont nous vérifierons le réalisme. En particulier, cela nous amènera à remettre en cause le choix d'un modèle paramétrique de type exponentiel par les auteurs de l'ouvrage déjà cité, choix qui ne s'appuyait sur aucune vérification et que nos résultats semblent contredire. Finalement, nous mobiliserons cette modélisation des processus de mise en couple pour étudier l'influence du fait d'être à l'école sur le choix d'un conjoint de même niveau d'études.

I Choix de la classification des diplômes

Le choix d'une nomenclature de diplômes est essentiel à la mesure de nombreux phénomènes, mais c'est plus particulièrement le cas pour l'étude de l'homogamie : celle-ci se définit en effet par rapport à un groupe. Le plus haut diplôme obtenu n'est donc pas seulement une variable dont on va tester l'effet sur d'autres variables : c'est ce qui définit le phénomène même que nous voulons identifier. D'où l'intérêt de choisir avec soin le ou les regroupements.

Le travers assez naturel dans ce type de démarche consiste à regrouper assez mécaniquement les diplômes inférieurs, considérés comme relativement identiques du point de vue de l'observateur, tout en conservant une plus grande finesse aux niveaux les plus élevés. Cette tentation est assez dangereuse concernant l'enquête Biographies et entourage, étant donné qu'elle traite de générations anciennes qui possèdent en majorité des diplômes peu élevés au regard des standards contemporains, sans parler des conceptions du chercheur en sciences sociales, qui ont toutes les chances d'être déformées en accord avec sa position sociale.

Il est d'autant plus nécessaire de garder ce problème à l'esprit que les nomenclatures contemporaines qui font référence, comme celle utilisée par l'INSEE pour l'enquête Emploi, sont bien plus précises dans le haut que dans le bas de la hiérarchie des diplômes. Le tableau 26 illustre clairement ce phénomène : la moitié la moins diplômée de la population est décrite par seulement trois catégories (aucun diplôme, certificat d'études, brevet d'études du premier cycle ou brevet d'études primaires supérieures), alors que le quart supérieur est décrit par pas moins de six catégories. Si cette asymétrie peut légitimement refléter le fonctionnement institutionnel du système scolaire, qui accorde des titres plus distincts ou distingués aux parcours les plus longs, elle peut aussi être source de distorsions dans les mesures statistiques, la plus évidente étant une sous-estimation mécanique de l'homogamie d'éducation pour les classes supérieures, qui sont considérées par la nomenclature comme plusieurs groupes différents[32]. De manière générale, l'utilisation d'une nomenclature fine à une extrémité de l'échelle et grossière de l'autre tendra, ainsi que nous l'avons déjà remarqué, à rendre visibles des variations relativement subtiles d'un côté tout en en masquant de l'autre.

32 Ce problème est, on l'a dit, statistiquement résolu par le recours à l'*odds ratio*, mais la technique de mesure ne peut pas pour autant résoudre ce que le chercheur ne veut pas voir, à savoir que des groupes distincts sur le papier n'en forment peut-être, à une échelle d'analyse donnée, qu'un seul.

	% par diplôme		% cumulé	
	Emploi 2001	**Biographies et entourage**	**Emploi 2001**	**Biographies et entourage**
Aucun	**20,9**	14,8	**21**	15
CEP	18,6	**23,8**	**40**	39
BEPC/BEPS	12,1	**13,5**	52	52
CAP/BEP	**15,2**	10,7	**67**	63
BEI, BEC, BEA...	1,2	**1,4**	**68**	64
Bac pro/technique Brevet professionnel	**2,3**	1,3	**70**	66
Bac général	6,9	**10,0**	**77**	76
Paramédical	**2,0**	0,9	**79**	76
BTS/DUT	**3,7**	3,2	**83**	80
1er cycle	1,8	**2,2**	**85**	82
2e et 3e cycles	9,7	**13,8**	94	**96**
Grande école	**5,7**	4,4	100	100
Total	100,0	100,0	-	-

Tableau 26 : Proportion des individus par diplôme, comparaison avec l'enquête Emploi

Champ : individus nés entre 1930 et 1950 et résidant en Île de France en 2000-2001

Pour l'enquête Emploi, l'échantillon est de 5428 individus (aucune valeur manquante). Pour Biographies et entourage, des 2830 individus de l'enquête (27 valeurs manquantes ou hors catégories ont été ignorées).

Les pondérations sont réalisées, respectivement, à l'aide des variables EXTRI et POIDFIN.

Les chiffres en gras indiquent l'enquête pour laquelle le pourcentage est le plus élevé.

	Effectifs	% brut	% après pondération
Aucun	325	11,6	14,7
CEP	619	22,1	23,8
BEPC/BEPS	397	14,2	13,5
CAP	222	7,9	8,2
CFA	1	0,0	0,0
BEP	71	2,5	2,4
BEI, BEC, BEA, BEH, BES	47	1,7	1,4
BT/Bac professionnel	16	0,6	0,5
Bac technologique	27	1,0	0,7
Bac général	322	11,5	10,0
BTS/DUT	97	3,5	3,2
1er cycle	67	2,4	2,2
Paramédical	32	1,1	0,9
2e cycle	208	7,4	6,6
Professionnel supérieur	70	2,5	2,3
3e cycle/professeurs	221	7,9	7,3
Grande école	62	2,2	2,1
Total	2804	100,0	100,0

Tableau 27 : Proportion d'individus par diplôme (nomenclature détaillée)

Champ : individus nés entre 1930 et 1950 et résidant en Île de France en 2000-2001

La pondération est réalisée à l'aide de la variable POIDFIN.

26 valeurs manquantes ont été ignorées.

La nécessité de réaliser des regroupements au sein de la nomenclature se fait pourtant clairement sentir (Tableau 27). Outre le fait que certaines catégories ont des effectifs trop faibles pour autoriser des traitements statistiques significatifs, il ne fait sens d'étudier l'homogamie qu'à une échelle relativement agrégée : il est assez anecdotique de savoir si les (peu nombreux) titulaires d'un brevet de technicien choisissent leur conjoint exactement dans la même catégorie de diplôme ; bien plus intéressant est de savoir, par exemple, s'il existe une homogamie parmi les titulaires des diplômes les plus élevés de l'enseignement professionnel.

Nous avons donc procédé à des regroupements sur la base de deux critères : d'une part, les catégories retenues pour constituer notre nomenclature de référence devaient diviser la population en groupes de tailles sensiblement équivalentes, en l'absence de raison d'agir de manière contraire ; d'autre part, les catégories ainsi fusionnées devaient apparaître assez proches en termes de position dans l'espace matrimonial pour que ce regroupement fasse sens. Notre principal outil pour en juger est une analyse des correspondances réalisée sur la table

d'homogamie des premiers couples[33], qui permet de juger de la proximité des catégories de diplômes sans aucun présupposé autre que le comportement matrimonial des groupes considérés.

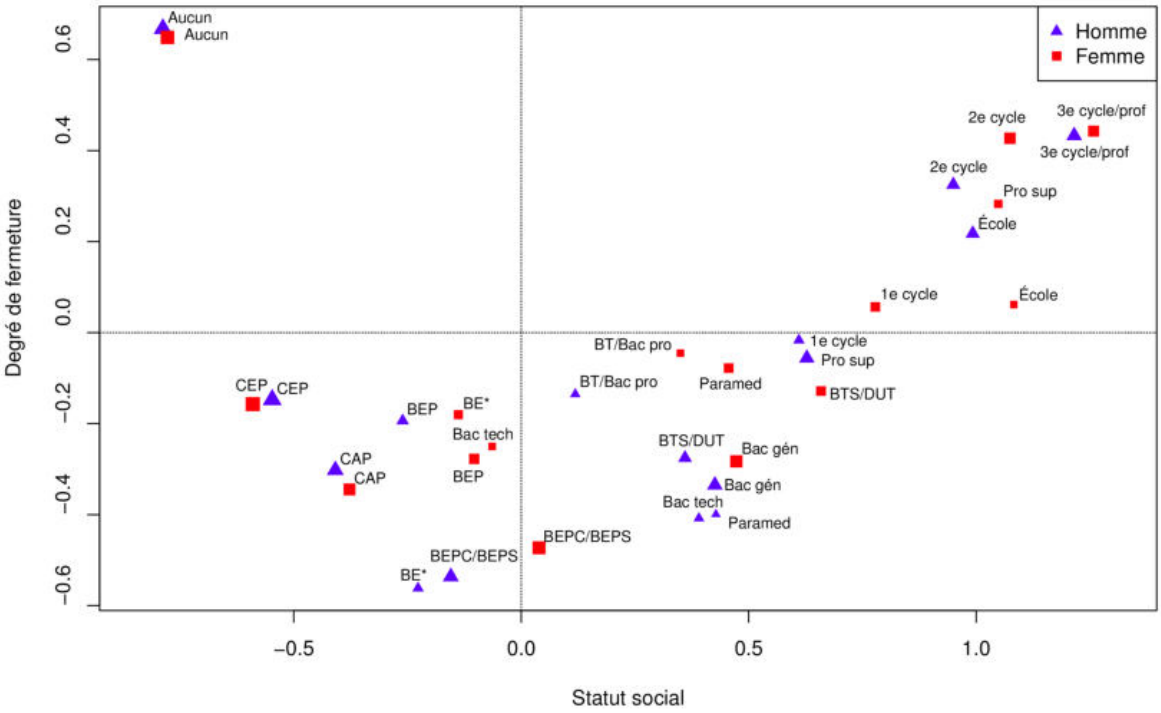

Figure 13 : L'espace matrimonial des premiers conjoints selon le diplôme

Champ : individus nés entre 1930 et 1950 et résidant en Île de rance en 2000-2001

L'échantillon est de 2505 couples (217 valeurs manquantes ou hors catégories ont été ignorées, et 108 individus n'ont déclaré aucun couple dans la base).

La pondération est réalisée à l'aide de la variable POIDFIN.

Les deux premiers axes rendent comptent respectivement de 57 % (axe horizontal) et 20 % (axe vertical) de la variance totale.

L'analyse des correspondances met en évidence quatre groupes assez bien séparés : les individus sans aucun diplôme, ceux avec un diplôme inférieur au baccalauréat, ceux disposant du baccalauréat ou d'un diplôme du premier cycle du supérieur d'une durée inférieure ou égale à deux ans, et enfin ceux ayant suivi des études supérieures plus longues. Les quelques catégories qui semblent atténuer ces découpages sont en fait très faibles en effectifs : les catégories Brevet de technicien/Bac professionnel (0,5 % des effectifs) et Bac technologique (0,7 %) changent de groupe selon que l'on considère les hommes ou les femmes. À part ces anomalies, les groupes semblent stables, ce qui autorise à procéder à des regroupements sans trop d'hésitations.

Dans la classification retenue, les groupes disposant d'effectifs déjà assez élevés ont été conservés, avec comme ordre de grandeur la taille du groupe le plus important, celui des titulaires du certificat d'études. Les autres ont été fusionnés avec les groupes les plus proches :

33 Le choix des premiers couples plutôt que des unions d'un ordre supérieur est arbitraire, mais, les secondes unions étant largement minoritaires, cette décision n'affecte que marginalement les résultats.

- Les individus sans diplôme, les titulaires du seul certificat d'études ou du BEPC ont conservé leurs groupes respectifs, du fait de leur taille.
- Les groupes CFA (une personne), BEP (2,4 %) et BE* (BEI, BEC, BEA, etc., 1,4 %) ont été intégrés au groupe CAP. Si le groupe BE* semble relativement éloigné, ses très faibles effectifs expliquent potentiellement la position erratique des points correspondant aux hommes et aux femmes de cette catégorie.
- La catégorie brevet de technicien/baccalauréat professionnel et la catégorie baccalauréat technologique ont été fusionnées, du fait principalement de l'inversion de positions, déjà notée plus haut, qu'ils connaissent selon le sexe, et qui interdit de les classer dans l'un des groupes adjacents. Leur mise à part permet d'éviter de fausser la composition des autres groupes.
- Les catégories BTS/DUT (0,9 %), premier cycle universitaire (3,2 %) et paramédical (2,2 %) ont été regroupées du fait de leur très nette proximité. Le baccalauréat général a été gardé à part à cause de ses effectifs déjà importants.
- Le dernier groupe, celui des études universitaires longues a été divisé en deux : deuxième cycle (6,6 %), enseignement professionnel supérieur (2,3 %) et grandes écoles (2,1 %) d'un côté ; troisième cycle et concours de professeur de l'autre (7,3 %), grâce aux effectifs assez importants et à la séparation assez marquée de ce dernier groupe.

En %	**Hommes**	**Femmes**	**Ensemble**
Aucun	13,0	10,3	11,6
CEP	21,3	22,7	22,1
BEPC/BEPS	11,3	16,7	14,2
CAP/BEP/BE*	12,7	11,6	12,2
BT/Bac professionnel/technologique	2,0	1,2	1,5
Bac général	9,7	13,1	11,5
1er cycle	10,5	8,6	9,5
2e cycle/Grande école	10,6	8,8	9,6
3e cycle/professeur	8,9	7,0	7,9
Total	100,0	100,0	100,0

Tableau 28 : Proportion d'individus par diplôme et par sexe (nomenclature agrégée)

Champ : individus nés entre 1930 et 1950 et résidant en Île de France en 2000-2001

Aucune pondération n'est appliquée ici, en cohérence avec les analyses qui suivent.

Catégorie	Codes inclus	Description
Aucun	0, 9	Aucun diplôme, sans objet
Valeur manquante	88, 99, 8, 20	Non réponse, ne sait pas, autres
CEP	1	Certificat d'études et Diplôme de fin d'études obligatoire
BEPC/BEPS	2	Brevet d'études du premier cycle, Brevet d'études primaires supérieures, Brevet élémentaire
CAP/BEP/BE*	11, 12, 13, 14	Certificat d'aptitude professionnelle Brevet d'études professionnelles Brevet d'enseignement industriel, commercial, secrétaire, comptabilité, agricole, social, hôtelier
Bac pro/technologique, Brevet de technicien	15, 16	Baccalauréat professionnel et technologique, Brevet de technicien
Bac général	3	Baccalauréat général, première partie de baccalauréat général, Brevet supérieur
1er cycle, BTS/DUT, Paramédical	4, 17, 18	Diplôme universitaire généraliste d'une durée inférieure à deux ans Brevet de technicien supérieur, Diplôme universitaire de technologie Diplôme d'infirmière, sage femme, kinésithérapeute, manipulateur radio, assistante sociale, éducateur spécialisé
2e cycle, grande école, professionnel supérieur	5, 7, 19	Diplôme universitaire de deuxième cycle Diplôme de grande école (général ou professionnel)
3e cycle/professeurs	6	Diplôme universitaire de troisième cycle, concours de professeur

Tableau 29 : Récapitulatif du regroupement des diplômes retenu

II Traitement des données

Nous disposons à travers Biographies et entourage d'informations concernant 2830 personnes et de tous leurs conjoints successifs jusqu'à l'âge de 50 ans au minimum. Les couples retenus dans l'enquête sont ceux ayant duré au moins un an, mais sans autre exigence : la cohabitation n'est pas nécessaire, non plus que le mariage. Nous ne nous intéressons ici qu'aux premières unions des individus, pour des raisons de simplicité et d'homogénéité ; même si les individus qui sont interrogés ont pu se mettre en couple pour la première fois après l'enquête, s'intéresser aux premiers couples nous permet aussi d'être sûr de disposer d'informations presque exhaustives, tant les probabilités de première mise en couple sont faibles au-delà de 45 ans (*cf.* ci-dessous).

Dans un premier temps, nous avons préparé les données en réalisant un appariement entre les individus interrogés (dénommés dans la suite Ego), qui sont représentatifs de la population d'Île de France[34], et leurs premiers conjoints, grâce à un identifiant de questionnaire et à un

34 Nous n'appliquons dans ce qui suit aucune pondération, puisque nous nous intéressons plus à des associations entre variables qu'à des proportions au sein de la population de référence. De plus, la comparaison en termes

numéro d'ordre des conjoints. Pour cela, nous avons créé une table contenant, le cas échant, le premier couple de chaque Ego, les informations concernant Ego et son premier conjoint, ainsi que l'âge de mise en couple. Les individus n'ayant eu aucun conjoint correspondent à des observations censurées, et on prend comme point de censure l'âge à la date de l'enquête. Dans les autres cas, l'âge à la première mise en couple est retrouvé, en fonction des cas, en prenant la première date mentionnée : date de première cohabitation (dans la plupart des cas), date de mariage si la cohabitation lui est ultérieure, et enfin date de début de relation en l'absence de cohabitation ou de mariage[35].

Grâce à une question posée en début de questionnaire, on peut confronter le nombre de conjoints retrouvés dans la base au nombre de conjoints que les individus ont initialement déclaré avoir eu : 9 personnes ont déclaré initialement un conjoint de moins que ceux qu'ils ont finalement décrits dans la suite ; 11 personnes ont fait l'inverse ; 10 autres ont déclaré deux conjoints de plus que ceux qu'ils ont décrits, et un autre en a déclaré 6 et n'en a décrit finalement qu'un. Ceci peut s'expliquer par des erreurs de codage, mais surtout par le fait que, d'un côté, des conjoints éphémères et trop lointains ont pu être négligés (la fiche conjoint est assez détaillée pour rebuter l'interrogé s'il a initialement compté dans sa déclaration un conjoint presque tombé dans l'oubli), et de l'autre par le fait que la démarche longitudinale peut amener les interrogés à parler de conjoints qu'ils avaient spontanément exclus de leur décompte. Finalement, ces anomalies ne représentant que 1,1 % de l'échantillon, elles ne nous semblent pas poser véritablement problème, surtout si l'on considère qu'un premier conjoint ne représente pas la même chose dans la vie d'un individu qui a oublié de le mentionner, que dans celle d'un autre qui y a pensé. Le profil de ces oublis serait certainement intéressant à étudier, mais dans le cadre d'une enquête d'une taille bien supérieure.

Finalement, on retrouve un premier couple pour 2722 individus : 105 personnes n'en ont déclaré aucun à aucun endroit du questionnaire ; 3 en ont déclaré initialement, mais ne les ont pas décrits dans la suite. 3,7 % des individus de l'échantillon n'ont donc eu aucun conjoint au moins jusqu'à l'âge de 50 ans. Parmi les individus dont on a retrouvé un premier couple, 24 n'ont donné aucune des trois dates possibles de début d'union (que l'on pourrait théoriquement retrouver de manière grossière en se référant aux événements mentionnés avant et après) : on dispose donc au total de couples datés pour 2698 individus sur 2830.

	Échantillon	**Couples déclarés**	**Couples dans la base**	**Couples datés**
Inclus	2 830 (100 %)	2 725 (96,3 %)	2 722 (96,2 %)	2 698 (95,3 %)
Exclus	0 (0 %)	105 (3,7 %)	105+3=108 (3,8 %)	105+3+24=132 (4,7 %)

Tableau 30 : Effectifs à chaque étape du traitement

de diplôme entre l'échantillon et l'enquête Emploi 2001 restreinte à la population concernée (tableau 26, p. 81) a montré que les écarts restaient raisonnables, ce qui garantit que les mécanismes observés dans l'échantillon ne s'éloignent pas trop de ceux valables dans cette dernière.

35 La rigueur aurait plutôt incliné à prendre dans les trois cas la date de début de relation, mais d'une part cette information n'est pas disponible dans l'enquête dans le cas où les individus ont cohabité ou se sont mariés plus tard, et d'autre part il semble difficile de définir une date de début de relation qui fasse sens de manière identique pour tous les interrogés, qui plus est cinquante ans après les faits. Quoi qu'il en soit, la faiblesse des effectifs concernés (la cohabitation et le mariage étant évidemment la règle) rend ce problème assez secondaire en pratique.

III Les rythmes de mise en couple

Approche non paramétrique

Une fois les données préparées, commençons par étudier sans *a priori* le comportement matrimonial des individus de notre échantillon. Nous utilisons en première approche l'estimateur de Kaplan-Meier séparément pour les hommes et pour les femmes[36] (Figure 14). Cette technique vise simplement à calculer le taux d'individus n'ayant jamais vécu en couple à chaque âge, en tenant compte des individus pour lesquels on n'a observé aucun couple à la date de l'enquête (censure à droite) ; dans Biographies et entourages, cette censure n'intervient jamais avant 50 ans. Il permet aussi de calculer des intervalles de confiance, ce qui peut s'avérer utile sur des échantillons de petite taille ; néanmoins, l'intervalle de confiance à 95 % étant ici très étroit, on ne l'a pas représenté sur la figure.

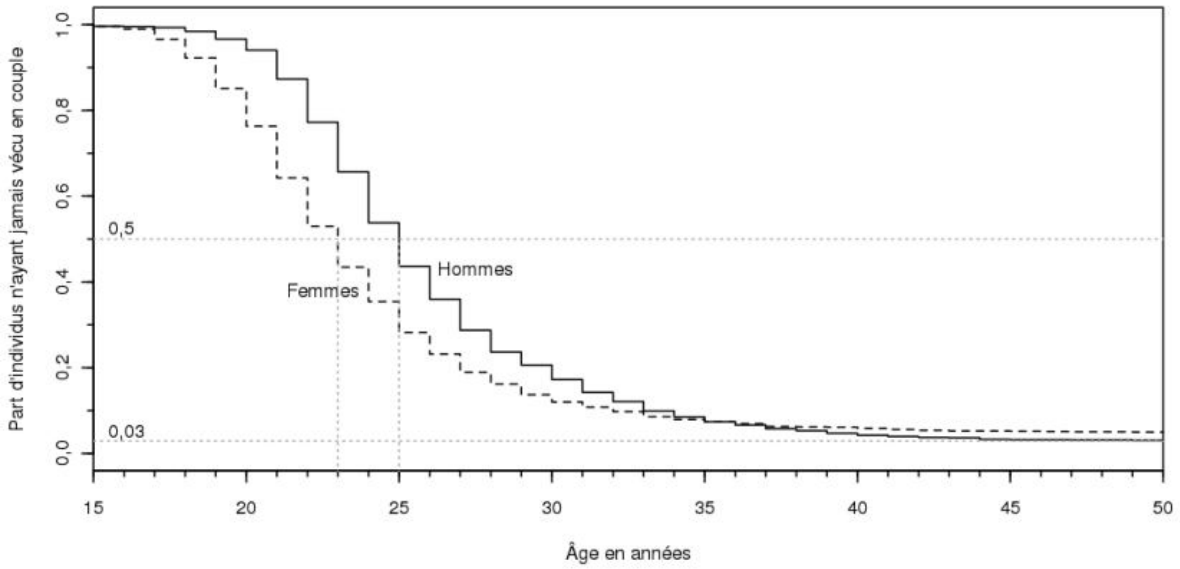

Figure 14 : Rythme de première mise en couple (estimateur de Kaplan-Meier)

Note : On ne considère ici ni les ruptures, ni les remises en couple. Ainsi, la proportion mesurée ici ne correspond pas à celle des personnes vivant en couple à un âge donné, des séparations ayant pu intervenir entre-temps.

Cette méthode nous permet de calculer l'âge médian de première mise en couple : 23 ans pour les femmes et 25 ans pour les hommes. On peut aussi observer une asymptote horizontale à droite qui signale une stabilisation du taux d'individus n'ayant jamais vécu en couple autour de 4,5 % pour les femmes, et 2,9 % pour les hommes (écarts-types des

36 Même s'il aurait pu être pratique, pour des questions d'effectifs, de créer deux entrées pour chaque couple en inversant à chaque fois la position de l'homme et de la femme, cette méthode aurait posé à nos yeux deux problèmes : d'une part ce sont les Ego qui sont représentatifs, et non leurs (premiers) conjoints ; d'autre part, l'inversion des positions aurait conduit à une symétrisation forcée des couples, alors que les individus prennent en partie des conjoints hors des cohortes représentées dans l'échantillon (condition qui permet que l'âge moyen à la première mise en couple diffère selon le sexe au sein de l'échantillon).

estimateurs de 0,58 % et de 0,46 %). Ce décalage, assez net et significatif au seuil de 10 %, illustre un phénomène plus général : alors que les femmes sont bien plus précoces quant à la mise en couple, les parts respectives des hommes et des femmes n'ayant jamais vécu en couple deviennent égales à 35 ans, et ce sont finalement les hommes qui restent le moins souvent définitivement célibataires.

En %	**Aucun**	**CEP**	**BEPC/ BEPS**	**CAP/ BEP/BE**	**BT/Bac pro/tech**	**Bac général**	**1er cycle**	**2e cycle/ École**	**3e cycle/ prof.**
Hommes	**4,0**	3,5	4,0	1,8	0,0	3,1	1,4	3,6	1,7
Femmes	2,0	**4,5**	**5,3**	**4,7**	**5,9**	**4,1**	**4,7**	**4,6**	**8,7**

Tableau 31 : Taux d'individus n'ayant déclaré aucun couple selon le sexe et le diplôme

On considère ici les individus n'ayant pas déclaré avoir déjà été en couple à la date de l'enquête, à laquelle ils avaient entre 50 et 70 ans. On a mis en gras le taux le plus élevé de chaque colonne, et souligné celui le plus élevé de chaque ligne.

Modèles semi-paramétriques

L'estimateur de Kaplan-Meier présente l'avantage de ne poser aucune hypothèse sur les rythmes de mise en couple et sur l'effet exercé sur ceux-ci par les variables explicatives (ici, le sexe). Cependant, cette approche rencontre rapidement ses limites lorsque l'on introduit des variables possédant de nombreuses modalités : dans notre cas, étudier l'influence du diplôme exigerait de représenter deux courbes (hommes et femmes) pour chaque niveau de diplôme retenu, ce qui devient rapidement difficile à interpréter et statistiquement non significatif. En outre, l'étude de l'influence de facteurs variant avec le temps, comme le fait d'être à l'école une année donnée, est impossible. Il est donc utile de recourir à un modèle plus précisément spécifié, et plus puissant à condition que les hypothèses qui le sous-tendent soient respectées.

Ainsi, il est d'usage de recourir tout d'abord à des modèles semi-paramétriques comme le modèle à « probabilités proportionnelles » de Cox. Celui-ci tente de rendre compte de l'effet des variables explicatives sur les probabilités de mise en couple tout en évitant de contraindre la forme de la courbe de base, qui mesure les probabilités instantanées de mise en couple indépendamment des variables explicatives. Afin de réduire les multiples courbes générées par une méthode non paramétrique à un seul coefficient, exprimant l'effet de la modalité *à tout âge*, le modèle de Cox pose une hypothèse assez restrictive et qui doit être testée : que les variables explicatives ne font que multiplier les probabilités de mise en couple par un certain facteur, indépendamment de l'âge des individus. Pour cela, les courbes de survie des différents groupes constitués sur la base des variables utilisées doivent être parallèles.

C'est cette hypothèse qu'il nous faut vérifier avant de pouvoir utiliser ce modèle. Dans notre cas, nous ne considérons pas le sexe comme une variable explicative classique : puisqu'il est clair que le sexe modifie profondément les rythmes de mise en couple, de manière non multiplicative (les courbes se croisent autour de 35 ans), l'hypothèse n'est pas vérifiée. Nous faisons donc du sexe une variable de strate, c'est-à-dire que nous autorisons le modèle à utiliser une courbe de probabilité de base différente pour chaque sexe. En revanche, puisque c'est l'effet du diplôme que nous souhaitons étudier grâce au modèle de Cox, nous introduisons ce dernier facteur comme une variable classique[37].

37 Il peut paraître étrange de recourir à une mesure statique du niveau d'éducation, le diplôme, alors que l'on étudie des processus fondamentalement chronologiques. S'il est vrai que la variable diplôme est souvent

Avant même de regarder les coefficients de régression obtenus par le modèle, nous testons l'hypothèse de proportionnalité des probabilités. Le résultat est sans appel (Tableau 32) : sur les huit niveaux de diplôme utilisés, seulement deux satisfont cette hypothèse, et elle est rejetée au niveau global avec la certitude maximale (*p-value* inférieure à 10^{-16}). Ce résultat n'est pas complètement inattendu, si l'on considère que les mêmes mécanismes sociaux jouent probablement, au moins en partie, pour le sexe et pour le diplôme. Intuitivement, si les groupes de diplôme ont tendance à se mettre en couple de manière accélérée immédiatement après la fin des études, la proportionnalité des probabilités ne sera pas respectée, puisqu'on n'assistera pas à une multiplication, mais plutôt à une forme de décalage des courbes de probabilité.

Diplôme	Rho	Chi-2	p-value
Aucun	*Réf.*	*Réf.*	*Réf.*
CEP	0,01	0,5	4,80E-01
BEPC/BEPS	0,06	9,7	1,90E-03
CAP/BEP/BE*	0,05	5,8	1,60E-02
BT/Bac pro/tech	0,03	1,9	1,60E-01
Bac général	0,09	21,0	5,00E-06
1er cycle	0,10	29,0	5,80E-08
2e cycle/École	0,12	39,0	3,30E-10
3e cycle/professeur	0,11	35,0	4,10E-09
Global		96,0	0,00E+00

Tableau 32 : Test de l'hypothèse de proportionnalité de l'effet du diplôme

Afin de mieux comprendre comment l'hypothèse de proportionnalité des probabilités est violée, on représente les résidus de Schoenfeld calculés pour chaque individu selon son âge de première mise en couple (Figure 15, pour certains diplôme seulement). Pour que l'hypothèse soit respectée, ces résidus ne doivent pas varier de manière significative avec l'âge, et doivent donc suivre une ligne horizontale. Dans notre cas, on observe que les lignes représentant la tendance moyenne des résidus varient assez nettement sur l'axe vertical, et surtout que les marges d'erreurs de ces lignes (en pointillés) sont extrêmement faibles, signe que toute variation est significative. En particulier, on peut noter que le modèle sous-estime l'augmentation des chances de mise en couple pour les titulaires d'un CEP aux âges inférieurs, et les surestime ensuite entre 22 et 24 ans par contrecoup ; les phénomène est

anachronique, puisque les individus peuvent se mettre en couple avant de l'obtenir, la solution qui consiste à compter, à chaque période, le nombre d'années d'éducation déjà suivies est non seulement peu pratique (elle exige de posséder une information que peu d'enquêtes fournissent, ou de faire des hypothèses fortes sur une progression scolaire « normale »), mais surtout assez hypocrite. L'utilisation du seul nombre d'années de scolarité entretient en effet le mythe selon lequel, à chaque étape de la scolarité, chaque individu aurait la même probabilité de la poursuivre ou non, n'étant pas différencié des autres jusqu'à ce qu'il quitte le système scolaire. Le diplôme finalement obtenu en dit bien plus long sur les origines de l'individu, sur les ambitions qu'il nourrit et qui influencent directement ses plans de mise en couple, et finalement sur un univers culturel que, l'ayant exclu de l'analyse, l'on aurait beau jeu de qualifier par la suite d'« hétérogénéité inobservée ». Ainsi, si le déroulement chronologique de l'éducation semble utile à l'analyse, on pourra toujours l'introduire *en sus* sous la forme dite plus haut, ce qui présentera l'avantage de ne mesurer qu'un temps passé à l'école, net des déterminants déjà captés par le diplôme.

moins accentué pour les titulaires d'un BEPC ou d'un BEPS. De manière générale, on observe une sous-estimation aux âges inférieurs, qui s'accroît légèrement avec le niveau du diplôme, et une très légère sous-estimation aux âges avancés.

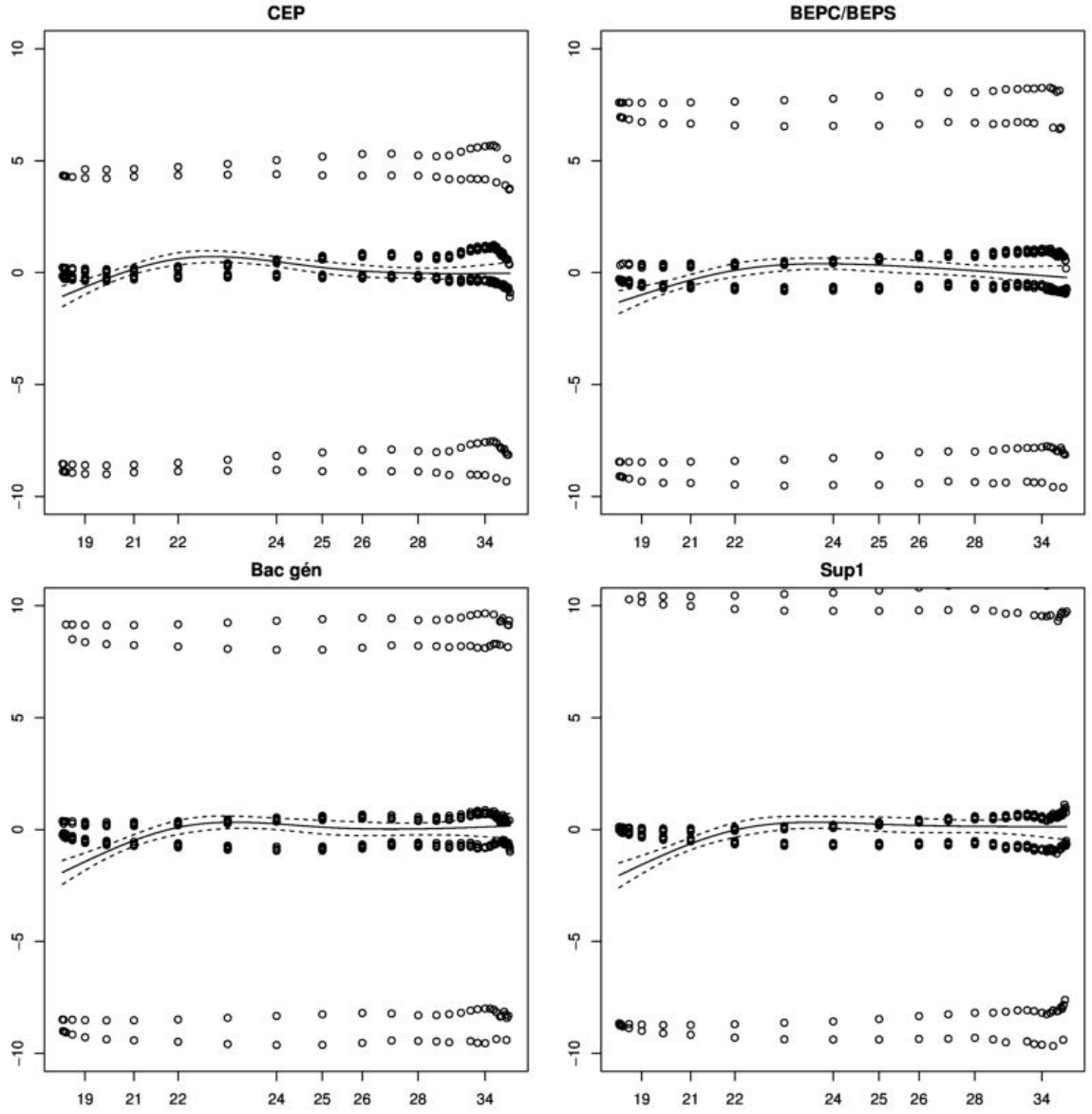

Figure 15 : Résidus de Schoenfeld concernant l'effet de quelques niveaux de diplôme

Il est souvent recommandé de découper l'échelle des âges en intervalles sur lesquels l'effet des variables respecterait l'hypothèse de proportionnalité des chances, et d'introduire ces intervalles comme variable d'interaction avec la variable explicative problématique (ici, le diplôme). Dans notre cas, cependant, cette technique n'a pas suffi à résoudre le problème, même en utilisant un nombre ridiculement élevé d'intervalles – ce que laissaient attendre les très mauvais résultats du test et la forme assez irrégulière des courbes. Nous nous tournons donc vers une autre famille de modèles.

Modèles paramétriques

Étant donné que la modélisation semi-paramétrique a échoué du fait du non respect de l'hypothèse de proportionnalité des probabilités de mise en couple, nous ne pouvons *a fortiori* pas recourir à des modèles paramétriques qui se fondent sur cette hypothèse. Il ne nous reste donc qu'à tester l'adéquation à nos données de la seconde famille de modèles paramétriques : celle des modèles à temps de sortie accéléré. Ceux-ci postulent que les variables explicatives affectent la distribution des chances non pas en la multipliant uniformément, mais en multipliant l'échelle temporelle par un certain coefficient. Du point de vue sociologique, ces modèles mesurent donc la plus ou moins grande précocité des individus, ce qui correspond *a priori* mieux à notre cas, de manière intuitive du moins.

Que nous utilisions des modèles paramétriques à temps de sortie accéléré, ou à probabilités proportionnelles, il est tout d'abord nécessaire de modéliser la courbe de distribution sur laquelle agiront les variables explicatives (qui peuvent en fait dans un premier temps être absentes). On dispose pour cela de plusieurs fonctions de distribution standard d'une complexité croissante (*cf.* Figure 16, p. 93, pour une illustration). La plus simple, la distribution exponentielle, suppose que les individus encore célibataires ont à chaque période des chances identiques de se mettre en couple : elle ne correspond clairement pas à nos données, selon lesquelles cette probabilité croît très fortement à partir de 15 ans, et décroît progressivement après 25 ans. Les autres fonctions varient en complexité et en souplesse, la règle étant de choisir la distribution la plus simple qui décrit correctement les données. Pour cela, on se fie principalement à l'indicateur AIC (critère d'information d'Akaike), qui juge de la qualité du modèle en fonction de sa fidélité aux données et de sa parcimonie (nombre de degrés de liberté) ; on recourt aussi à titre de vérification, à l'indicateur BIC (critère d'information bayésien), qui tient compte de la taille de l'échantillon utilisé.

Pour commencer, nous introduisons une seule variable explicative, la plus déterminante : le sexe. Cela nous permettra d'examiner si une distribution correspond bien à nos données, et si l'hypothèse du temps de sortie accéléré s'accorde au moins à ce critère. Avant de lancer les modèles, on retire tous les individus censurés, c'est-à-dire ceux qui ne s'étaient jamais mis en couple à la date de l'enquête, à laquelle ils avaient entre 50 et 70 ans. Cette étape est nécessaire puisque les fonctions testées varient entre 0 et 1, n'autorisant pas une partie des individus à ne jamais connaître l'événement étudié. Cela n'est pas un problème, puisque notre analyse peut se diviser en deux étapes : tout d'abord, pour chaque catégorie étudiée, on observe le taux de mise en couple définitif (selon la définition introduite plus haut), puis, pour ceux qui se sont mis en couple, on observe la fonction de distribution estimée par le modèle.

Nous utilisons le paquet rms (*Regression modeling strategies*) et sa fonction psm (*Parametric survival model*), qui présentent l'avantage de fournir des méthodes simplifiant la représentation des distributions et des résidus (*cf.* Figure 16), ainsi que des distributions prédites pour un individu doté de caractéristiques données – opérations qui doivent être réalisées manuellement avec les autres paquets disponibles (survival, sur lequel rms se fonde, ou eha, qui sera utilisé plus bas pour d'autres opérations).

Distribution	Déviance	Δ déviance	Degrés de liberté	AIC	BIC
Exponentielle	19 067	0	2	19 071	19 083
Gaussienne	16 216	-2 851	3	16 222	16 161
De Weibull	16 137	-79	3	16 143	16 240
Log-normale	15 561	-576	3	15 567	15 584
Log-logistique	**15 395**	-166	3	**15 401**	**15 419**

Tableau 33 : Comparaison de l'adéquation des différentes distributions à nos données

AIC = déviance + 2 × (nombre de degrés de liberté)

BIC = déviance + ln(N_{obs} = 2 698) × (nombre de degrés de liberté)

On a mis en gras les valeurs minimum de la déviance, de l'AIC et du BIC.

Notons que les critères AIC et BIC ordonnent globalement les modèles de la même manière, à part une inversion des places des distributions gaussienne et de Weibull, qui se joue de peu. Comme prévu, la distribution exponentielle est la moins adaptée à nos données. Les autres distributions sont de plus en plus fidèles à mesure que leur complexité augmente. Finalement, les fonctions log-normale et log-logistique sont sensiblement identiques quant à la qualité de leur prédiction, avec un léger avantage pour la distribution log-logistique, que nous retiendrons donc.

Quatre de ces distributions sont représentées graphiquement ci-dessous (Figure 16) : on peut vérifier que la distribution exponentielle est clairement inadéquate. La distribution de Weibull est déjà bien meilleure, mais elle ne rend pas correctement compte de la rupture de rythme qui s'observe dès que les individus commencent à se mettre en couple (autour de 20 ans), et, inversement, du ralentissement qui s'observe après 25 ans. Au contraire, la courbe log-logistique semble s'adapter parfaitement aux données, et la courbe log-normale n'en est pas non plus très loin. Nous pouvons donc considérer que nous disposons d'une (très) bonne modélisation de la courbe générale de mise en couple.

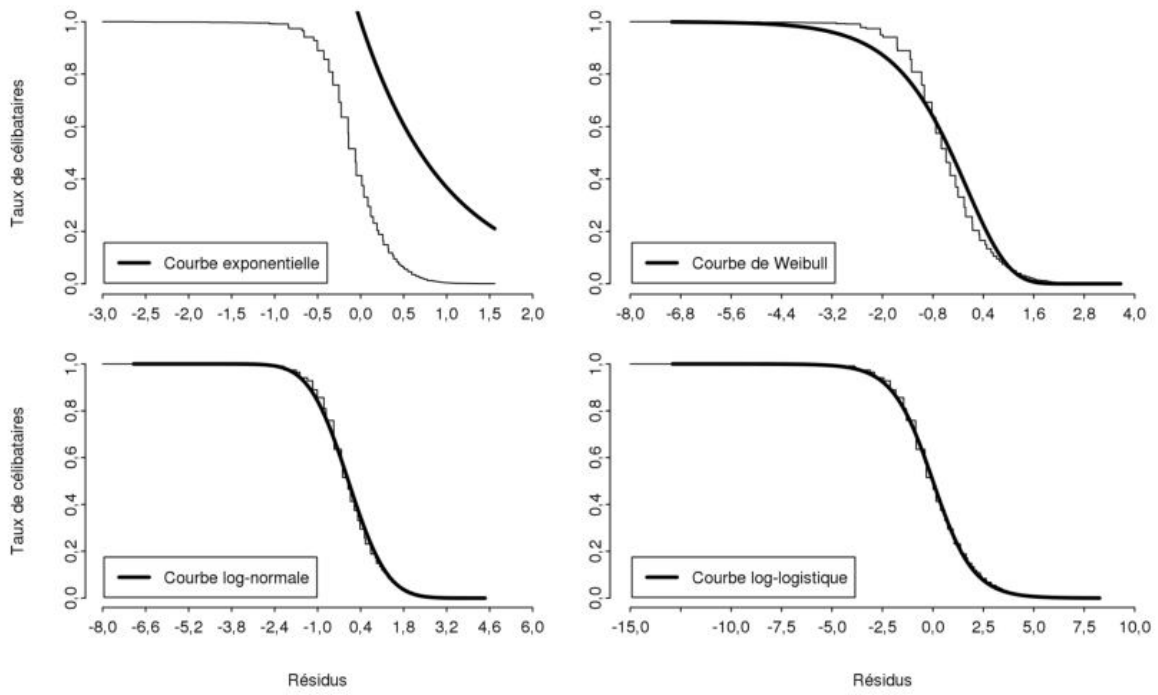

Figure 16 : Test des différentes lois de distribution

Le résidu correspond à l'écart normalisé entre l'âge de mise en couple de l'individu et l'âge moyen de mise en couple de la loi théorique.

Qu'en est-il cependant de la différence entre hommes et femmes ? Le paquet rms nous permet de vérifier que l'hypothèse d'accélération temporelle est respectée, en représentant les résidus de Cox-Snell pour chaque sexe (Figure 17, premier graphique). Ceux-ci correspondent à l'écart entre la fonction de séjour prédite par le modèle et celle observée dans les données, telle que calculée par l'estimateur de Kaplan-Meier (qui est le complément à 1 de la fonction de survie présentée en figure 14, page 87). Le graphique nous permet de remarquer que la distribution log-logistique a tendance à sous-estimer le taux de personnes en couple : c'est particulièrement le cas pour les femmes jeunes (lorsque le taux d'individus en couple est faible) et pour les hommes âgés (lorsque ce taux est élevé). Ce point signale que la transformation de l'échelle temporelle ne suffit pas à expliquer complètement la différence de comportement entre hommes et femmes. Mais globalement, cette distribution estime bien les données : les écarts entre les courbes sont relativement faibles, d'autant qu'on n'observe pas de déviation systématique entre les taux observés et prédits qui affecte un sexe et pas l'autre.

D'après le modèle log-logistique, les femmes ont donc un profil de mise en couple accéléré de 18 % par rapport aux hommes : le coefficient mesurant l'effet d'être une femme plutôt qu'un homme est de -0,2 en échelle logarithmique, soit 0,82 en échelle naturelle (±0,02 avec un risque d'erreur de 5 %). Cette valeur est relativement robuste, puisque même avec les modèles exponentiel et de Weibull, qui correspondaient très mal aux données, elle est estimée respectivement à 0,84 et 0,88 (avec une marge d'erreur cependant plus importante).

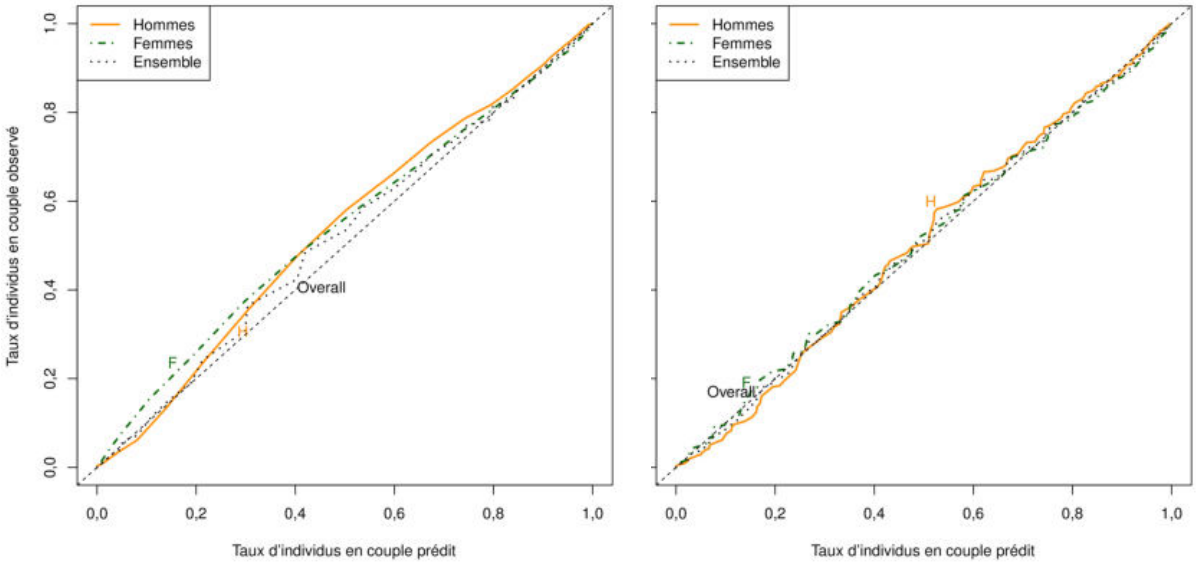

Figure 17 : Résidus de Cox-Snell pour le modèle log-logistique, sans puis avec le diplôme

Puisque nous disposons maintenant d'un modèle raisonnable des rythmes de mise en couple selon le sexe, nous pouvons procéder à une complexification du modèle en ajoutant le diplôme des individus, variable sans doute essentielle dans la détermination de l'âge de mise en couple.

Alors que cet ajout visait surtout à étudier l'effet du diplôme sur la mise en couple, il a un effet très positif sur la variable sexe : une fois tenu compte du diplôme, le modèle s'ajuste nettement mieux aux profils de mise en couple respectifs des hommes et des femmes, une grande partie des déviations précédemment observées étant maintenant capturées par l'effet du diplôme (Figure 17, second graphique). Les indicateurs AIC et BIC montrent tous deux que l'ajout du diplôme au modèle l'améliore de manière importante (le premier descend à 15 095, le second à 15 161, soit deux variations hautement significatives).

En outre, les différentes catégories de diplôme s'ajustent très bien au modèle log-logistique. Ce point est particulièrement intéressant : il pourrait signaler le fait que cette distribution décrit véritablement un processus social sous-jacent à la mise en couple, mesurant ainsi les probabilités individuelles de mise en couple, et pas seulement celles valables au niveau agrégé. Même s'il est difficile de confirmer cette hypothèse, on peut noter que la loi log-logistique correspond à des modèles théoriques de diffusion, dans lesquels l'influence de l'entourage incite l'individu à adopter un comportement ou un produit, ce qui n'est pas dénué de sens dans le cas de la mise en couple (Dieckmann, 1990). Mais d'autres modèles, qui se fondent sur des hypothèses comportementales quant à la recherche d'un conjoint et aux préférences des individus, aboutissent à des distributions sensiblement proches de la distribution log-logistique ; sans les avoir testées, il est impossible de conclure.

Alors que l'ajustement au modèle est la plupart du temps très bon, on doit néanmoins noter que deux catégories – les individus sans diplôme et ceux disposant au plus d'un diplôme professionnel de niveau baccalauréat – dévient nettement de la distribution log-logistique, de manière assez spectaculaire (Figure 18). Cette déviation semble s'expliquer dans les deux cas par une nette divergence de comportements entre hommes et femmes au sein de cette catégorie de diplôme. Ainsi, les femmes non diplômées se mettent en couple nettement plus vite que le modèle ne le prévoit, alors que c'est l'inverse pour les hommes. En ce qui concerne les titulaires d'un diplôme équivalent au baccalauréat professionnel, les femmes de ce groupe

se mettent en couple plus rapidement que la prédiction à des âges très jeunes, puis rejoignent la prédiction, au moment où les hommes de cette catégorie commencent à se mettre plus rapidement en couple qu'attendu.

Il est délicat d'analyser les comportements de ces deux groupes à partir de ces seuls graphiques, puisqu'ils représentent une déviation, non par rapport à la moyenne de la population, mais par rapport à la courbe log-logistique correspondant le mieux au comportement de cette population, à laquelle sont appliqués les deux coefficients d'accélération du temps qui permettent le mieux de décrire, respectivement, le comportement de chaque sexe et de chaque groupe de diplôme. Les écarts observés sont donc mesurés par rapport à ce qui serait attendu d'un individu qui croise ces deux caractéristiques de sexe et de diplôme : on mesure en fait l'écart d'un comportement à une norme théorique dans laquelle l'individu serait en cohérence avec les deux groupes d'appartenance introduits dans le modèle, mais sans que l'association spécifique de ces deux appartenances n'entraîne l'apparition d'un effet propre.

Si ces déviations sont considérées comme gênantes, il est facile de les éliminer en introduisant des effets d'interaction entre sexe et diplôme, éventuellement pour les seules catégories qui ne suivent pas le modèle attendu. Nous ne le ferons pas ici puisque notre attention ne se porte pas spécialement sur ces groupes problématiques, et que les déviations observées restent, pour notre étude, d'ampleur raisonnable. Notons aussi que les effectifs du groupe des titulaires d'un diplôme professionnel de niveau baccalauréat sont très faibles (43 personnes, soit 1,5 % de l'échantillon), ce qui laisse soupçonner que cette originalité ne soit qu'un artefact. Il serait cependant intéressant d'étudier plus précisément les comportement inattendus du groupe des non diplômés, pour lesquels le fait d'être un homme ou une femme transforme radicalement l'effet du diplôme. On sait par exemple que les hommes peu diplômés et les femmes très diplômées ont plus de mal à se mettre en couple que les autres (*cf.* tableau 31, page 88), à cause de la norme sociale qui incite les hommes à choisir une femme moins éduquée qu'eux, et inversement. Ce phénomène mériterait d'être étudié de manière longitudinale, mais cela constituerait véritablement une analyse à part.

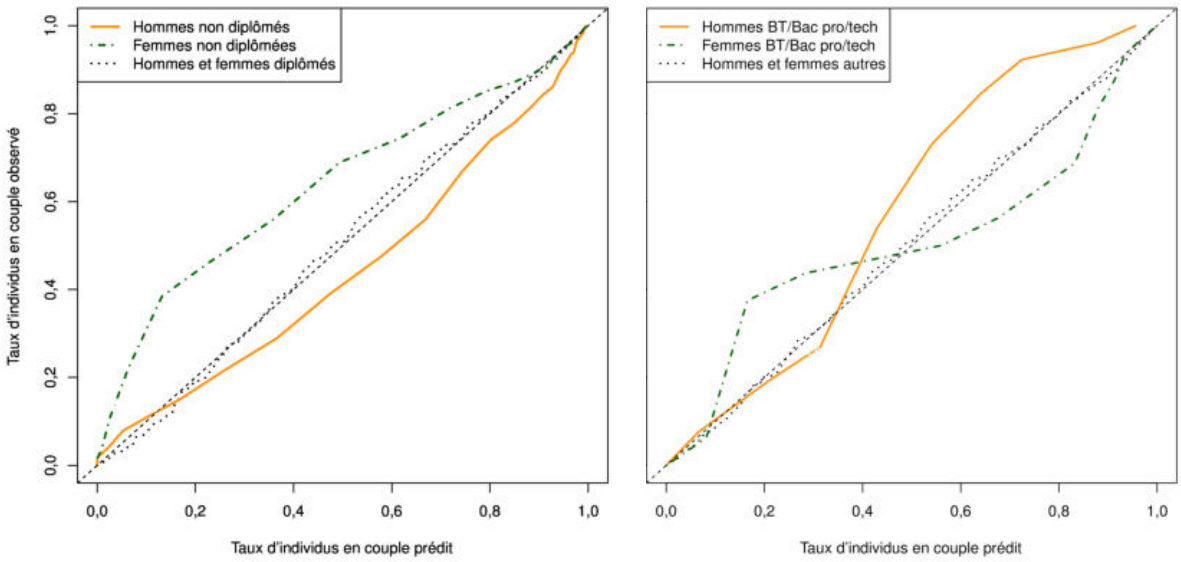

Figure 18 : Résidus de Cox-Snell du modèle log-logistique pour les deux groupes problématiques

Les coefficients de vitesse de mise en couple calculés par le modèle sont reportés ci-dessous (tableau 34). À la lecture des intervalles de confiance, on peut considérer que deux groupes différents se détachent : les individus dont le diplôme le plus élevé est un diplôme professionnel, ou un diplôme général inférieur au BEPS/BEPC se mettent en couple plus ou aussi vite que la catégorie de référence ; à l'inverse, ceux qui ont un diplôme supérieur au BEPS/BEPC se mettent en couple plus lentement. Avec les effectifs dont nous disposons, il est difficile de réaliser des distinctions plus fines : notamment, la distinction entre les diplômés du second et du troisième cycle du supérieur, et ceux du premier cycle (ou du baccalauréat général) ne tient pas à un seuil de confiance de 10 % (Figure 19).

	Coefficient	**Écart-type**	**Exp(coefficient)**	**Variation par rapport à la référence (en %)**
Femme	-0,196***	(0,013)	0,82	-18
Aucun	-0,045	(0,026)	0,96	-4
CEP	-0,079***	(0,022)	0,92	-8
CAP/BEP/BE*	-0,038	(0,025)	0,96	-4
BEPS/BEPC	*Réf.*	-	*Réf.*	*Réf.*
BT/Bac pro/tech	-0,096	(0,053)	0,91	-9
Bac général	0,070***	(0,026)	1,07	+7
1er cycle	0,082***	(0,027)	1,09	+9
2e cycle/École	0,151***	(0,027)	1,16	+16
3e cycle/professeur	0,153***	(0,029)	1,17	+17
Constante	2,504	(0,021)	-	-
Échelle de la distribution	-1,650	(0,016)	-	-

Tableau 34 : Coefficients d'accélération du modèle log-logistique

Champ : individus ayant déclaré au moins un couple.

** Significatif au seuil de 5 %, ** de 1 %, *** de 0,1 %.*

L'échantillon utilisé est de 2 672 individus. Modèle à 9+2 degrés de liberté.

Déviance : 15 074, AIC : 15 095, BIC : 15 161.

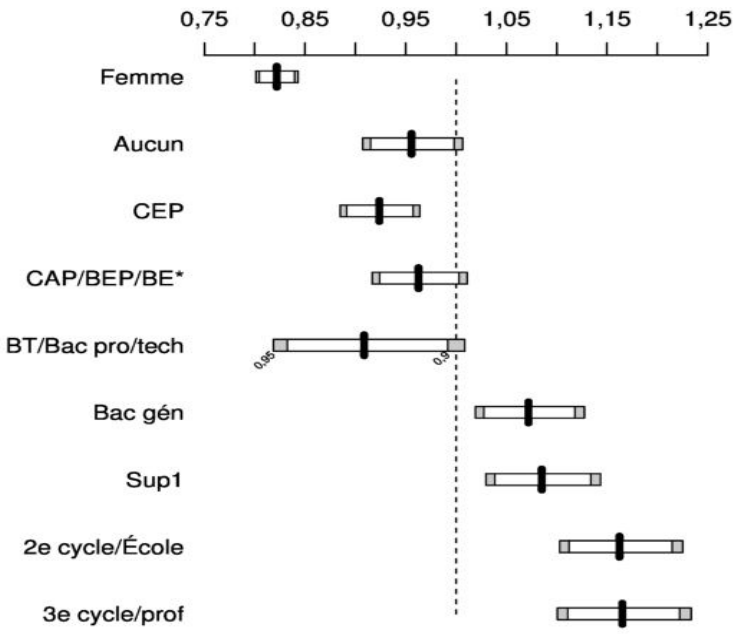

Figure 19 : Coefficients d'accélération du modèle log-logistique

On a pris pour référence les hommes pour la variable sexe, et les titulaires d'un BEPS/BEPC pour le diplôme (en raison de la position moyenne de cette catégorie). Les barres horizontales représentent les intervalles de confiance à 90 % (blanc) et à 95 % (gris).

Lecture : Les individus dont le plus haut diplôme est un certificat d'études primaires (CEP) se mettent en couple 8 % plus rapidement que ceux dont le plus haut diplôme est un brevet d'enseignement primaire supérieur (BEPS) ou un brevet d'études de premier cycle du second degré (BEPC). Au contraire, les titulaires d'un diplôme universitaire du troisième cycle ou d'un concours de professeur se mettent en couple 17 % plus lentement que ces derniers.

Finalement, la figure 20 représente les probabilités instantanées de mise en couple de trois niveaux de diplôme, ainsi qu'elles sont prédites par le modèle log-logistique.

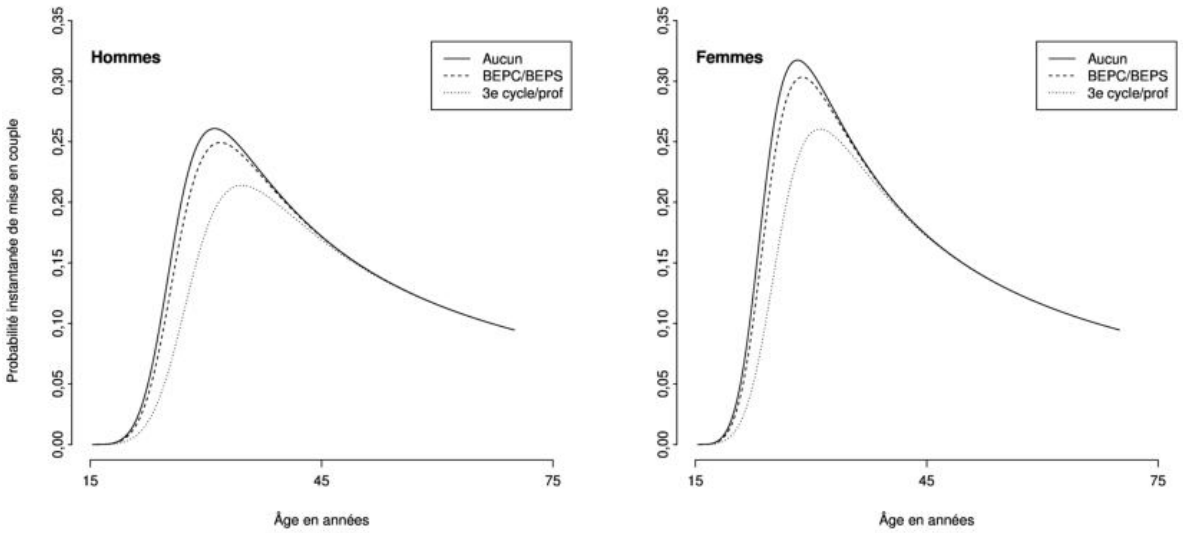

Figure 20 : Taux instantanés de mise en couple prédits par le modèle log-logistique

IV Homogamie et mise en couple

Nous pouvons maintenant mesurer l'influence du moment de mise en couple sur l'homogamie. Nous avons construit une variable reflétant le fait qu'un Ego est homogame, hypogame ou hypergame, selon que son diplôme est respectivement égal, supérieur ou inférieur à celui de son conjoint. Nous disposons aussi de l'âge de première activité d'Ego, qui correspond donc à la fin des études. Pour étudier la mise en couple homogame ou non, nous mettons en place un modèle tenant compte du type de couple formé.

Notre objectif sera ici de mesurer l'influence de la scolarisation sur le type de couple formé, défini par les trois états décrits ci-dessus. Il ne nous est pour cela plus possible d'utiliser la fonction psm, puisqu'elle ne prend pas en charge les variables évoluant au cours du temps, ce qui est ici le cas du fait ou non d'être à l'école une année donnée. Nous nous tournons donc vers le paquet eha (*Event history analysis*) et sa fonction aftreg (*Accelerated failure time regression*), qui ont pour particularité de retourner des coefficients dont le signe est inversé par rapport à ceux de psm : un coefficient positif signale donc dans la suite une accélération du temps, donc une probabilité de mise en couple plus élevée. Auparavant, nous transformons nos données de manière à ne plus avoir une ligne par individu, mais une ligne par année observée pour chaque individu jusqu'à sa première mise en couple (format dit « long »). Pour chaque année, on note si l'individu poursuit ou non ses études, et si cette année est ou non celle de sa première mise en couple.

Le premier modèle (tableau 35) est identique à celui présenté plus haut (tableau 34, p. 96), et n'est inclus qu'à titre de vérification : à part l'inversion du signe des coefficients, les résultats sont bien les mêmes (ils sont ici repris sous leur forme logarithmique seulement). Le second modèle tient compte, en plus du sexe et du diplôme, du type de couple formé. Cette information est essentielle à la description des rythmes de mise en couple, puisque le modèle est très nettement amélioré d'après les indicateurs AIC et BIC[38]. On observe que les individus ayant choisi un conjoint de niveau d'études supérieur se mettent en couple légèrement plus tard que les autres (les homogames – la référence –, mais aussi les hypogames), à sexe et diplôme fixés. Ce phénomène est assez naturel, pour autant que l'on se rappelle qu'il faut être deux pour constituer un couple : même si nous ne tenons compte ici que des caractéristiques d'Ego, celles du conjoint influencent aussi le rythme de la mise en couple. Dans le même sens que nous l'observons pour Ego, un diplôme plus élevé du conjoint retarde mécaniquement la mise en couple, et inversement. Nous n'approfondirons pas ici cette voie, mais il serait utile d'étudier l'influence du niveau de diplôme du conjoint, même si nous en avons déjà un aperçu grâce à la combinaison du niveau d'études d'Ego et du type de couple qu'il a formé.

38 De manière rigoureuse, la comparaison des indicateurs AIC et BIC doit être réalisée sur des modèles aux échantillons identiques, ce qui n'est pas le cas des modèles 1 et 2 présentés ici. Précisons donc que l'application du modèle 1 à l'échantillon restreint du modèle 2 donne un AIC de 14 050 et un BIC de 14 114, soit deux valeurs nettement supérieures à celles obtenues pour le modèle 2.

	Modèle 1		Modèle 2	
	Coef.	Écart-type	Coef.	Écart-type
Homme	*Réf.*		*Réf.*	
Femme	0,196***	(0,013)	0,204***	(0,013)
Aucun	0,045	(0,026)	0,044	(0,028)
CEP	0,079***	(0,022)	0,078***	(0,022)
CAP/BEP/BE*	0,038	(0,025)	0,026	(0,026)
BEPS/BEPC	*Réf.*		*Réf.*	
BT/Bac pro/tech	0,096	(0,053)	0,081	(0,054)
Bac général	-0,069**	(0,026)	-0,083**	(0,027)
1er cycle	-0,082**	(0,027)	-0,098***	(0,028)
2e cycle/École	-0,150***	(0,027)	-0,175***	(0,028)
3e cycle/professeur	-0,153***	(0,029)	-0,191***	(0,031)
Hypogame			0,014	(0,017)
Homogame			*Réf.*	
Hypergame			-0,064***	(0,016)
Forme de la distribution	2,549	(0,019)	2,523	(0,022)
Échelle de la distribution	1,649	(0,016)	1,665	(0,017)
Individus	2 672		2 493	
Épisodes (années observées)	33 755		31 569	
Déviance	15 074		14 004	
Degrés de liberté	9+2		11+2	
AIC	15 096		14 030	
BIC	15 160		14 105	

Tableau 35 : Modèles appliqués au format « long », variables fixes

Champ : individus ayant déclaré au moins un couple.

** Significatif au seuil de 5 %, ** de 1 %, *** de 0,1 %.*

Le BIC est calculé sur la base du nombre d'événements inclus dans le modèle, c'est-à-dire le nombre d'individus sans valeurs manquantes (tous les individus retenus s'étant ici mis en couple).

L'hypothèse selon laquelle les couples homogames, hypogames et hypergames ont chacun une distribution différente semble devoir être rejetée : pour ce modèle alternatif, l'AIC est de 19 538 et le BIC de 19 649, soit des valeurs assez nettement supérieures à celles des deux modèles ci-dessus. Nous sommes donc bien en présence d'une modification de l'échelle temporelle, plus que d'un comportement radicalement différent.

Enfin, on introduit dans le troisième modèle (tableau 36) le fait ou non de poursuivre ses études une année donnée. On ajoute aussi directement un effet différencié selon le type de couple formé (interaction entre type de couple et scolarisation), puisque cela améliore le modèle selon les deux indicateurs AIC et BIC, et constitue le point qui nous intéresse le plus

(seul le modèle incluant cette interaction est repris ici). Il apparaît ainsi que le fait d'être encore à l'école diminue fortement les probabilités de se mettre pour la première fois en couple : ne plus être à l'école accélère le rythme de mise en couple d'un facteur 5 (inverse de l'exponentielle d'un coefficient égal à -1,624), soit un effet approximativement triple de celui d'être une femme plutôt qu'un homme.

Plus intéressant encore, le fait d'être à l'école diminue le rythme de mise en couple plus fortement pour les individus hypogames et hypergames que pour les individus homogames ; ou, exprimé dans l'autre sens, le fait d'être sorti de l'école diminue les chances relatives de se mettre en couple avec une personne de même niveau d'études. Ce résultat démontre donc l'intuition selon laquelle le choix d'un conjoint pendant les études mène à plus d'homogamie. S'il est absolument clair que cet effet est significatif, son ampleur semble en revanche assez faible : on observe une augmentation du rythme de mise en couple de 20 % pour les hypergames par rapport aux homogames, et de légèrement moins pour les hypogames. On peut calculer (en additionnant les coefficients) que, par rapport à un individu ayant formé un couple du même type hors de l'école, le ralentissement du rythme de la mise en couple dû à la scolarisation est respectivement d'un facteur de 5,07, 5,46 et 6,11 pour les couples homogames, hypogames et hypergames.

Enfin, on peut remarquer l'inversion générale des coefficients mesurant le rythme de mise en couple selon le diplôme : une fois tenu compte du fait d'être sur les bancs de l'école une année donnée, ce sont les plus diplômés qui ont le rythme de mise en couple le plus élevé, alors que c'était le contraire en termes d'effet brut[39]. L'interprétation de ce phénomène est délicate : le fait que le niveau de diplôme net de la scolarisation présente un effet positif sur le rythme de mise en couple signale que les plus diplômés ne retardent pas leur mise en couple autant que le nombre d'années d'études qu'ils ont suivies le supposerait, si l'on se réfère au comportement des moins diplômés à cet égard. Ceci est relativement naturel, puisque les moins diplômés de notre échantillon quittent l'école à des âges encore assez jeunes pour se mettre en couple, alors que les plus diplômés n'ont, eux, aucune raison de retarder la mise en couple une fois qu'ils ont terminé leurs études – et ont même, l'âge croissant, des raisons de plus en plus fortes de trouver rapidement un conjoint. Ce phénomène peut aussi traduire une plus grande facilité des plus diplômés à se mettre en couple de manière générale, mais il est difficile de distinguer cet effet du « rattrapage » lié à la fin des études en l'état, notamment sans tenir compte de la durée écoulée depuis la fin des études.

39 On rappelle que le diplôme est ici une variable fixe, valable même pour les années durant lesquelles l'individu n'a pas encore terminé ses études (*cf.* note 37 p. 88).

	Modèle 3		Modèle 4	
	Coef.	Écart-type	Coef.	Écart-type
Homme	*Réf.*		*Réf.*	
Femme	0,401***	(0,006)	0,416***	(0,012)
Aucun	-0,203***	(0,012)	-0,199***	(0,012)
CEP	-0,110***	(0,010)	-0,100***	(0,010)
CAP/BEP/BE*	0,026**	(0,011)	0,018	(0,011)
BEPS/BEPC	*Réf.*		*Réf.*	
BT/Bac pro/tech	0,421***	(0,030)	0,451***	(0,028)
Bac général	0,200***	(0,014)	0,206***	(0,013)
1er cycle	0,304***	(0,014)	0,308***	(0,013)
2e cycle/École	0,392***	(0,014)	0,296***	(0,013)
3e cycle/professeur	0,450***	(0,017)	0,458***	(0,016)
Hypogame	-0,028**	(0,009)	0,020	(0,011)
Homogame	*Réf.*		*Réf.*	
Hypergame	-0,005	(0,008)	0,008	(0,011)
Sorti de l'école	*Réf.*		*Réf.*	
À l'école	-1,624***	(0,017)	-1,682***	(0,028)
Homogame × Sorti de l'école	*Réf.*		*Réf.*	
Hypogame × À l'école	-0,074**	(0,023)	-0,044	(0,035)
Hypergame × À l'école	-0,186***	(0,024)	-0,446***	(0,046)
Homme × Sorti de l'école			*Réf.*	
Femme × À l'école			0,079*	(0,034)
Homogame × Homme			*Réf.*	
Hypogame × Femme			-0,135***	(0,018)
Hypergame × Femme			-0,008	(0,017)
Homogame × Homme × Sorti de l'école			*Réf.*	
Hypogame × Femme × À l'école			-0,002	(0,046)
Hypergame × Femme × À l'école			0,338***	(0,053)
Forme de la distribution	1,997	(0,016)	2,093	(0,016)
Échelle de la distribution	3,637	(0,019)	3,605	(0,019)
Individus	2 493		2 493	
Épisodes (années observées)	31 569		31 569	
Déviance	136		-104	
Degrés de liberté	14+2		19+2	
AIC	169		-62	
BIC	262		60	

Tableau 36 : Modèles appliqués au format « long », variables dépendantes du temps

Cf. tableau 35 pour la lecture.

Les effets de la scolarisation ne semblent pas différenciés selon le diplôme obtenu : un modèle plus fin incluant des coefficients d'interaction entre scolarisation et diplôme n'apporte rien (aucun coefficient d'interaction n'étant significatif, AIC et BIC supérieurs à ceux du modèle 3). En revanche, ces effets sont bien différenciés selon le sexe (modèle 4)[40]. En introduisant l'interaction du second ordre entre sexe, type de couple formé et scolarisation, on améliore très nettement la description des données, et on retrouve un résultat classique selon lequel les femmes se mettent moins souvent en couple avec un homme de diplôme inférieur au leur. On observe aussi que les femmes se mettent un peu plus en couple lorsqu'elles sont encore à l'école que les hommes (accélération de 8 %). Mais surtout, on peut maintenant voir que les hommes se mettent beaucoup moins en couple avec une femme de diplôme supérieur au leur lorsqu'ils sont scolarisés (décélération de 56 %), soit plus que la différence de rythme générale entre hommes et femmes. En revanche, cet effet est très faible pour les femmes (de l'ordre de 10 %).

Au total, c'est donc plus l'hypergamie des hommes qui varie fortement en fonction de la scolarisation : ce phénomène n'est pleinement expliqué que lorsque l'interaction du second ordre est introduite, faisant apparaître un coefficient assez élevé[41]. Nous pouvons donc conclure que la mise en couple hypergame connaît un net clivage selon que l'intéressé est ou non scolarisé. Il est pourtant difficile d'expliquer ce phénomène en l'état des informations dont nous disposons, notamment sans considérer l'âge de la conjointe. En effet, puisque nous retenons le diplôme finalement obtenu, nous ne savons pas quel est son âge, et nous ignorons donc si, à la date de mise en couple, la femme avait déjà un diplôme supérieur à celui de l'homme, et si elle avait quitté l'école. Pour observer ce phénomène du côté des femmes, et pas seulement de celui des hommes, il faudrait alors introduire une variable reflétant le fait, pour la femme, d'avoir récemment quitté l'école – puisque le fait de ne plus être à l'école décrit une période bien trop étendue, qui dilue les effets (et qui ne peut donc servir que de référence).

Finalement, nous reproduisons la même analyse en recourant à trois modèles séparés, un pour chaque type de couple, afin d'obtenir des résultats directement comparables à ceux de Goux et Maurin[42] (tableau 37, page 104). Pour cette même raison, nous retenons maintenant comme modalité de référence le fait d'être encore scolarisé. On aboutit aux mêmes résultats que plus haut : le fait d'avoir quitté l'école augmente plus fortement les chances de se mettre en couple de manière hypergame, que de le faire de manière hypo- ou homogame, avec une différence d'environ 15 % (le ralentissement par rapport au individus non scolarisés ayant formé le même type de groupe est respectivement d'un facteur 6,10, 5,44 et 5,17). La différence, dans cette série de modèles, tient à ce que le coefficient mesurant cet effet parmi les individus hypogames n'est pas significativement différent de celui correspondant aux individus homogames, au contraire du modèle 3 vu plus haut.

40 Avec l'échantillon dont nous disposons, il n'est pas possible d'appliquer ces modèles séparément aux hommes et aux femmes, le système de variables étant exactement singulier (colinéarité) dans certains cas, et l'algorithme échouant à converger dans d'autres. Cela ne semble cependant pas strictement nécessaire puisque le modèle décrit bien, comme on l'a vu plus haut, les différences entre hommes et femmes, et que les termes d'interaction suffisent à fournir des informations détaillées sur les différences de comportements concernant les variables d'intérêt.

41 On a pu vérifier en recourant à plusieurs modèles intermédiaires que c'est bien l'ajout de l'interaction du second ordre, et non celui des trois interactions du premier ordre entre sexe, type de couple et scolarisation, qui change radicalement la valeur de ce coefficient.

42 Sans pouvoir, encore une fois, séparer hommes et femmes, cette opération nous amenant en fait à calculer six modèles séparés sur un échantillon qui est bien plus réduit que celui de l'enquête Formation et qualification professionnelle, que ces auteurs utilisaient (*cf.* note 40, p. 102).

En tout cas, nos résultats sont très exactement contraires à ceux obtenus par Goux et Maurin (tableaux pp. 74-75), qui observaient que le fait d'avoir quitté l'école augmentait plus fortement les probabilités de mise en couple homogame que celles de mise en couple hypo- et hypergame, ce qui les amenait à conclure que l'expansion scolaire n'avait pas joué dans le sens d'un renforcement de l'homogamie d'éducation en France. Ces résultats portaient cependant sur la population française en 1993 (individus nés entre 1934 et 1963), et sur les mariages uniquement, quand notre étude se concentre sur une génération née entre 1930 et 1950, dont la première mise en couple est un peu plus ancienne (et qui ne représente qu'une région française, même si on peut douter que les mécanismes diffèrent de ceux à l'œuvre dans le reste du pays).

Au niveau de la société dans son ensemble, et de l'impact de la scolarisation sur l'homogamie, il est important de souligner un paradoxe. Alors que nos recherches montrent que l'homogamie d'éducation (nette des variations de la structure de diplôme de la population) s'est affaiblie de manière sensible en France depuis les années 1960 (au moins au sens large, *cf.* la deuxième partie), si l'on suit les résultats obtenus ici, l'allongement de la scolarité aurait dû mécaniquement se traduire par un renforcement de cette même homogamie. L'ampleur de l'effet attendu reste certes modeste, mais son sens est bien clair. Il semble donc que deux hypothèses puissent expliquer ce phénomène. Premièrement, il est tout à fait possible que les comportements aient simplement changé depuis la génération étudiée dans Biographies et entourage. Deuxièmement, on peut aussi envisager que l'expansion éducative ait conduit à un effet complètement opposé à ce que l'on observait de manière statique sur une génération : rien n'oblige en effet à ce qu'une variation observée au niveau individuel (une scolarité plus ou moins longue entre individus partageant un contexte social identique) produise les mêmes résultats qu'une variation réalisée au niveau collectif (un allongement général de la scolarité qui transforme le contexte social dans lequel les individus évoluent). Cette seconde hypothèse rejoint en fait la première, en ce qu'elle peut constituer une explication de la transformation des comportements, elle-même liée à une transformation du rôle de l'éducation dans la constitution de l'espace social.

Ce résultat peut aussi éclairer la divergence, observée dans la deuxième partie, entre homogamie stricte et homogamie large d'éducation. Nous n'avons retenu ici qu'une définition stricte de l'homogamie, sans tenir compte de la distance entre les diplômes des conjoints observable au sein des couples hypo- et hypergames. Aussi, l'augmentation de l'homogamie au sens retenu dans cette dernière partie ne correspond qu'à celle de l'homogamie stricte : nos résultats sont donc congruents. En fait, une augmentation de la durée des études étant susceptible de provoquer, selon les résultats de notre étude longitudinale, à une augmentation de l'homogamie stricte, il est possible d'expliquer que, alors que toutes les dimensions de l'homogamie semblent s'affaiblir, cette même homogamie stricte d'éducation se renforce malgré tout. Si cette explication est juste, le phénomène que nous avons observé dans notre deuxième partie n'est que l'effet mécanique de l'allongement de la durée des études, et pas d'une plus grande préférence des individus pour une proximité en termes de diplôme.

	Hypogames		Homogames		Hypergames	
	Coef.	Écart-type	Coef.	Écart-type	Coef.	Écart-type
Homme	*Réf.*		*Réf.*		*Réf.*	
Femme	0,287***	(0,007)	0,428***	(0,013)	0,478***	(0,009)
Aucun	-		-0,326***	(0,025)	-0,110***	(0,013)
CEP	-0,074***	(0,012)	-0,214***	(0,023)	-0,031**	(0,011)
CAP/BEP/BE*	-0,016	(0,013)	-0,034	(0,025)	0,028	(0,019)
BEPS/BEPC	*Réf.*		*Réf.*		*Réf.*	
BT/Bac pro/tech	0,333***	(0,020)	0,557***	(0,068)	0,538***	(0,029)
Bac général	0,227***	(0,013)	0,102**	(0,033)	0,257***	(0,017)
1er cycle	0,258***	(0,017)	0,306***	(0,031)	0,269***	(0,025)
2e cycle/École	0,308***	(0,013)	0,104**	(0,031)	0,449***	(0,022)
3e cycle/professeur	0,478***	(0,014)	0,318***	(0,036)	-	
À l'école	*Réf.*		*Réf.*		*Réf.*	
Sorti de l'école	1,694***	(0,011)	1,643***	(0,024)	1,809***	(0,016)
Forme de la distribution	3,769	(0,040)	3,649	(0,044)	3,987	(0,043)
Échelle de la distribution	4,401	(0,035)	3,089	(0,033)	3,855	(0,032)
Individus	796		774		923	
Épisodes	10 284		9 555		11 730	
Déviance	-1 360		900		-541	
Degrés de liberté	9+2		10+2		9+2	
AIC	-1 337		924		-519	
BIC	-1 286		980		-466	

Tableau 37 : Modèles séparés par types de couple

Champ : individus ayant déclaré au moins un couple.

** Significatif au seuil de 5 %, ** de 1 %, *** de 0,1 %.*

Conclusion

Thème classique et aujourd'hui quelque peu négligé de la sociologie, l'homogamie reste pourtant un champ de recherches actif dans plusieurs pays, et dans lequel nous espérons avoir montré qu'il est possible et nécessaire de faire progresser nos connaissances grâce à de nouvelles approches. Facilement victimes de biais liés à la durée de vie variable des couples, ou de problèmes méthodologiques, les études récentes, et en particulier françaises, semblent avoir manqué un certain nombre de phénomènes. Le recours à plusieurs enquêtes successives nous a permis d'éliminer ces difficultés fondamentales, autorisant ainsi une mesure plus fiable des évolutions de l'homogamie d'éducation et de statut social sur les quarante dernières années.

D'autre part, l'utilisation systématique de différents types de modèles log-linéaires et log-multiplicatifs nous a fourni un moyen sûr d'étudier la manière dont l'homogamie s'est transformée, une fois tenu compte des évolutions structurelles de la population. Au total, notre démarche met en évidence une diminution généralisée de l'homogamie d'éducation aussi bien que socio-professionnelle, de l'ordre d'un tiers sur quarante ans pour la première, et de légèrement moins pour la seconde, une fois tenu compte des transformations structurelles de la population. Mais l'introduction de modèles originaux autorisant une variation séparée de la diagonale des tables d'homogamie a aussi rendu visible un mouvement passé jusqu'ici inaperçu : une divergence entre homogamie stricte (égalité des diplômes des conjoints) et homogamie large d'éducation (proximité des niveaux d'éducation), la première restant relativement stable malgré le net affaiblissement de la seconde.

Alors que les études portant sur l'Amérique du Nord (Schwartz et Mare, 2005 ; Mare, 2008 ; Hou et Myles, 2008) concluent à une augmentation de l'homogamie nette d'éducation, nous avons donc des raisons de penser que la France se situe sur une trajectoire différente : parmi les études françaises récentes (Forsé et Chauvel, 1995 ; Goux et Maurin, 2003 ; Vanderschelden, 2006a), seuls Goux et Maurin ont observé une telle augmentation, et nous avons souligné plus haut les limites méthodologique de cette partie de leur étude. Quant aux autres pays développés, trop peu de travaux nous semblent disponibles pour tirer des conclusions générales : outre l'ouvrage collectif de Blossfeld et Timm (2003b), qui conclut à une augmentation de l'homogamie d'éducation dans de nombreux pays européens, mais qui souffre des mêmes limites que le précédent, Raymo et Iwasawa (2005) observent pour le Japon un affaiblissement de cette même homogamie, et Katrňák (2008) une courbe en U pour trois anciennes démocraties populaires d'Europe de l'Est. La question de l'évolution de l'homogamie d'éducation dans les sociétés développées est donc à nos yeux loin d'être tranchée.

L'affaiblissement de l'homogamie socio-professionnelle que nous constatons est en accord avec une étude française (Vallet, 1986) , mais en contradiction avec l'autre (Vanderschelden, 2006a) ; notre étude et celle de Vallet portant sur des enquêtes successives, les résultats

différents obtenus par Vanderschelden découlent potentiellement du biais lié à l'utilisation d'une seule enquête rétrospective, du fait de la durée de vie variable des couples. Les études de cet aspect de l'homogamie portant sur d'autres pays sont rares : on ne recense qu'une publication (Verbakel, 2008), portant sur les Pays-Bas, et qui conclut à une stabilité de l'homogamie socio-professionnelle. Il est donc difficile de situer la France de ce point de vue. On peut remarquer que l'homogamie liée au statut socio-professionnel des parents (origine sociale) semble en diminution des deux côtés de l'Atlantique (Kalmijn, 1991 ; Forsé et Chauvel, 1995), résultat qui peut être rapproché de la diminution de l'homogamie liée au statut socio-professionnel des conjoints eux-mêmes ; ces deux phénomènes sont cependant bien distincts, et ne varient pas nécessairement dans le même sens. Sur ce point aussi, la sociologie de l'homogamie peut donc encore réaliser d'importants progrès.

Signe des tendances profondes de l'évolution des sociétés, l'homogamie nous fournit aussi un moyen privilégié d'observer la structuration à une période donnée de l'espace matrimonial, et donc de l'espace social en général. Cette démarche n'a, à notre connaissance, jamais été suivie, même si Desrosières (1978) s'en approchait, en recourant à la technique (un peu différente de nos modèles d'association lignes-colonnes) de l'analyse des correspondances, et pour une seule année. Nous avons ainsi pu identifier une transformation de l'espace matrimonial des diplômes : en 1969, l'échelle des diplômes se structurait selon une seule dimension face aux non diplômés, qui formaient clairement un groupe à part ; en 2009, cet espace oppose enseignement général et l'enseignement professionnel, d'une part, et les peu diplômés aux très diplômés, de l'autre. L'étude de l'homogamie nous offre donc un aperçu des distinctions scolaires et culturelles en vigueur dans un système éducatif et dans une société, phénomène difficilement objectivable par d'autres voies.

L'espace matrimonial des statuts socio-professionnels nous a en outre permis d'observer le rapprochement progressif sur quarante ans des indépendants et des agriculteurs de l'échelle des catégories salariées ; cette évolution fait du groupe des employés de bureau (IIIa dans la classification internationale Casmin) et des techniciens (V), qui était déjà au centre de l'échelle du salariat en 1969, le groupe pivot de la société. La classification française des PCS illustre ce phénomène de manière légèrement différente, les indépendants et les employés formant un vaste groupe central ouvert sans discrimination à la majorité de l'échelle sociale. Ces observations sont de nature à enrichir notre compréhension de la structure sociale, ainsi que le débat sur l'existence ou le retour des classes sociales, ou encore la question de la pertinence des nomenclatures couramment utilisées.

Enfin, point qui intéresse plus directement la question des processus de formation du couple, l'étude des espaces matrimoniaux nous a permis de vérifier que la norme d'hypergamie des femmes, selon laquelle le statut de l'homme doit être supérieur à celui de la femme au sein du couple, a disparu au cours des années 1970 et 1980. Cette transformation des comportements matrimoniaux s'observe aussi bien en ce qui concerne le diplôme que la catégorie socio-professionnelle (mesurée par la PCS). C'est donc ici le progrès de l'égalité des sexes telle qu'elle se traduit dans les comportements que l'étude de l'homogamie nette des transformations structurelles de la population nous permet de mesurer.

La mesure de l'homogamie et de ses évolutions a posé, et continue de poser de multiples problèmes. Mais, même une fois ces questions réglées, il est nécessaire de pousser plus avant les analyses afin de tenter de comprendre les processus qui y mènent, et de donner un sens aux phénomènes observés : c'est ce que nous avons ébauché dans notre troisième partie. L'analyse

longitudinale de la biographie des individus montre que la mise en couple au cours des études est bien un facteur de renforcement de l'homogamie d'éducation. Obtenu à partir de l'enquête Biographies et entourage de l'INED, ce résultat contredit celui de Goux et Maurin (2003), dont nous avons souligné certaines faiblesses.

Si l'on suit les conclusions établies dans notre troisième partie, l'allongement de la durée des études consécutif à la massification scolaire a naturellement conduit à une augmentation de l'homogamie d'éducation, au moins considérée dans son sens strict (égalité des diplômes des conjoints, plutôt que simple proximité des niveaux d'étude). Ce mouvement, conjugué à une diminution générale de l'homogamie au sens large (aussi valable pour le statut socio-professionnel), peut fournir une explication aux évolutions divergentes que nous observions en deuxième partie entre homogamie stricte et homogamie large d'éducation : la stabilité de l'homogamie stricte ne serait alors que la résultante de deux tendances contradictoires, l'une liée aux transformations des comportements, l'autre à celles du système éducatif. Cette hypothèse ne semble pas jusqu'ici avoir été l'objet de recherches.

La présente étude appelle, semble-t-il, plusieurs prolongements. Du point de vue des études transversales, il serait intéressant de renforcer nos résultats par l'introduction de plus d'années d'enquête : l'échantillon des enquêtes Emploi étant complètement renouvelé tous les trois ans jusqu'en 2003, et tous les 18 mois depuis, il est possible de mesurer de manière beaucoup plus fine le rythme des évolutions détectées ici, et de vérifier leur régularité. Ensuite, il paraît naturel d'associer à l'étude de l'homogamie d'éducation et de statut socio-professionnel celle de l'homogamie d'origine sociale, ce que les enquêtes Emploi rendent possible à partir de 1982 (pour un tiers de l'échantillon). L'articulation de ces trois types d'homogamie constitue aussi un point d'intérêt, afin par exemple d'examiner dans quelle mesure ils se recouvrent : les conjoints ont-ils tendance à partager une seule caractéristique commune, ou les proximités ont-elles au contraire tendance à faire boule de neige ? une partie des couples est-elle très homogame à plusieurs points de vue alors qu'une autre l'est très peu, ou l'homogamie est-elle un phénomène très également réparti ? La question des multiples dimensions de l'homogamie a rarement été abordée dans la littérature, bien qu'elle soit porteuse d'importants enjeux théoriques.

En ce qui concerne les études longitudinales, de plus amples investigations devront s'attacher à vérifier notre conclusion selon laquelle la scolarisation renforce l'homogamie d'éducation, notamment en tenant compte de l'homogamie au sens large et de la distance sociale, et pas seulement de l'homogamie stricte (conjoints de diplômes identiques). Il sera aussi nécessaire d'approfondir certaines découvertes, comme celle d'une extrême variation du rythme de mise en couple hypergame des hommes selon que l'intéressé est ou non scolarisé. L'étude de ce type de couple, généralement considéré dans la littérature comme « non traditionnel » (puisque c'est la femme qui possède alors le statut le plus élevé), est fortement susceptible d'éclairer les processus de formation du couple. Finalement, l'étude longitudinale d'autres dimensions de l'homogamie (notamment socio-professionnelle et d'origine sociale), et de l'impact sur celle-ci de la séparation et de la remise en couple est loin d'être dénuée d'intérêt. On voit que les analyses biographiques offrent de nombreux prolongements pour une compréhension plus fine des processus menant à l'homogamie.

L'enquête Biographies et entourage, que nous avons utilisée ici pour l'étude de franciliens nés entre 1930 et 1950, ne se limite pas à décrire une génération : à travers le récit que font les interrogés de la vie de leurs parents et de leurs enfants, il est possible de mesurer l'homogamie de générations nées au tournant du XXe siècle, pour lesquelles aucune enquête transversale n'existe (au prix néanmoins d'une perte de représentativité). Ce suivi sur plusieurs

générations permet aussi d'étudier l'influence de l'homogamie d'un couple sur celle de ses enfants, menant à la fermeture, ou au contraire à l'enchevêtrement des réseaux sociaux, ou à de potentiels phénomènes d'inertie (Mare et Schwartz, 2006 ; Mare, 2008).

La compréhension de l'homogamie dans un seul pays fournit des informations intéressantes sur la société et sur les mécanismes de choix du conjoint ; néanmoins, il est d'un grand intérêt de pouvoir situer une société dans le cadre plus général des évolutions suivies par d'autres pays similaires. On a déjà souligné que les études les plus fiables se sont concentrées sur une poignée de pays seulement, de sorte qu'il est difficile de se prononcer notamment sur le sens de l'évolution de l'homogamie parmi les pays d'Europe de l'Ouest. C'est pourquoi il nous semble utile d'étendre notre étude dans cette direction. D'un côté, les *Labour Force Surveys* européennes, équivalent des enquêtes Emploi françaises, constituent une source de données transversales susceptibles d'être utilisées en suivant une démarche similaire à celle développée dans notre seconde partie.

De l'autre, l'enquête SHARELIFE (*Survey of Health, Ageing and Retirement in Europe*, volet biographique), publiée récemment, fournit des informations longitudinales harmonisées grâce à un questionnaire commun à 12 pays européens. Conduite par l'Université de Tilburg grâce à un financement de la Commission européenne, cette enquête porte sur 19 000 personnes (dont, par exemple, 2 000 pour la France), appartenant sensiblement à la même génération que l'échantillon de Biographies et entourage, et contient, pour ce qui concerne notre recherche, le même type de renseignements. L'existence de cette enquête européenne permet d'éviter le recours à des enquêtes nationales aux modalités très différentes, sources de difficultés lors de la synthèse des résultats. Ces données offrent ainsi l'opportunité de comparer les processus de formation du couple dans une série de sociétés assez proches pour qu'une comparaison ait du sens, mais qui ont des raisons de différer de manière sensible.

Ainsi, si les techniques et les données – en particuliers longitudinales – ont longtemps fait défaut, elles ne constituent plus un obstacle à une étude sociologique rigoureuse et riche de l'homogamie : la voie est désormais ouverte.

Bibliographie

Agresti, Alan (**2002**), *Categorical Data Analysis*, New York, Wiley.

Audirac, Pierre-Alain (**1982**), « Cohabitation et mariage : qui vit avec qui ? », *Economie et statistique*, vol. 145 (1), pp. 41-59.

Becker, Gary S. (**1991**), « Assortative Mating in Marriage Markets », *in A Treatise on the Family*,

Blackwell, Deborah L. (**1998**), « Marital homogamy in the United States: The influence of individual and paternal education », *Social Science Research*, vol. 27, pp. 159–188.

Blackwell, Deborah L. et Daniel T. Lichter (**2004**), « Homogamy Among Dating, Cohabiting, and Married Couples », *The Sociological Quarterly*, vol. 45 (4), pp. 719-737.

Blau, Peter M. (**1967**), *Who Marries Whom?*, New York, Wiley.

Blossfeld, Hans-Peter (**2009**), « Educational Assortative Marriage in Comparative Perspective », *Annual Review of Sociology*, vol. 35, pp. 513-530.

Blossfeld, Hans-Peter, Sonja Drobnič et Götz Rohwer (**1998**), « Les trajectoires professionnelles des couples mariés en Allemagne. Une étude longitudinale de long terme de carrières des époux en Allemagne de l'Ouest », *Revue française de sociologie*, vol. 39 (2), pp. 305-351.

Blossfeld, Hans-Peter et Andreas Timm (**2003a**), « Who Marries Whom in West Germany? », *in Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, , pp. 18-35

Blossfeld, Hans-Peter et Andreas Timm (dir.) (**2003b**), *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht (Pays-Bas), Kluwer Academic Publishers.

Bourdieu, Pierre (**1972**), « Les stratégies matrimoniales dans le système de reproduction », *Annales. Histoire, Sciences Sociales*, vol. 27 (4/5), pp. 1105-1127.

Bourdieu, Pierre (**1979**), *La distinction. Critique sociale du jugement*, Paris, Minuit.

Bourdieu, Pierre (**1984**), « Espace social et genèse des « classes » », *Actes de la recherche en sciences sociales*, vol. 52 (1), pp. 3–14.

Bozon, Michel et François Héran (**1987**), « La découverte du conjoint: I. Évolution et morphologie des scènes de rencontre », *Population*, vol. 42 (6), pp. 943-985.

Bozon, Michel et François Héran (**1988**), « La découverte du conjoint: II. Les scènes de rencontre dans l'espace social », *Population*, vol. 43 (1), pp. 121-150.

Breen, Richard et Jan O. Jonsson (**2000**), « Analyzing educational careers: A multinomial transition model », *American Sociological Review*, vol. 65 (5), pp. 754–772.

Bumpass, Larry L. et James A. Sweet (**1972**), « Differentials in Marital Instability: 1970 », *American Sociological Review*, vol. 37 (6), pp. 754-766.

Chardon, Olivier et Thomas Amossé (**2006**), « Les travailleurs non qualifiés : une nouvelle classe sociale ? », *Economie et statistique*, vol. 393 (1), pp. 203-229.

Commaille, Jacques (**1974**), « Divorce, milieu social et situation de la femme », *Economie et statistique*, vol. 53 (1), pp. 3-21.

Desrosières, Alain (**1978**), « Marché matrimonial et structure des classes sociales », *Actes de la recherche en sciences sociales*, vol. 20 (1), pp. 97-107.

Desrosières, Alain et Laurent Thévenot (**2002**), *Les catégories socio-professionnelles*, Paris, La Découverte.

Desrosières, Alain, Alain Goy et Laurent Thévenot (**1983**), « L'identité sociale dans le travail statistique : la nouvelle nomenclature des professions et catégories socioprofessionnelles », *Economie et statistique*, vol. 152 (1), pp. 55-81.

Dieckmann, Andreas (**1990**), « Diffusion and Survival Models for the Process of Entry into Marriage », *in* Karl Ulrich Mayer et Nancy Brandon Tuma (dir.), *Event History Analysis in Life Course Research*, , pp. 170-183

Erikson, Robert et John H. Goldthorpe (**1992**), *The Constant Flux: A Study of Class Mobility in Industrial Societies*, Oxford, Clarendon Press.

Forsé, Michel et Louis Chauvel (**1995**), « L'évolution de l'homogamie en France. Une méthode pour comparer les diagonalités de plusieurs tables », *Revue française de sociologie*, vol. 36 (1), pp. 123-142.

Gilula, Zvi (**1984**), « On Some Similarities Between Canonical Correlation Models and Latent Class Models for Two-Way Contingency Tables », *Biometrika*, vol. 71 (3), pp. 523-529.

Girard, Alain (**1964**), *Le Choix du conjoint : une enquête psycho-sociologique en France*, Paris, PUF.

Goodman, Leo A. (**1979**), « Simple Models for the Analysis of Association in Cross-Classifications having Ordered Categories », *Journal of the American Statistical Association*, vol. 74 (367), pp. 537-552.

Goodman, Leo A. (**1986**), « Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-Linear Models Approach in the Analysis of Contingency Tables », *International Statistical Review*, vol. 54 (3), pp. 243-270.

Goodman, Leo A. et Michael Hout (**1998**), « Statistical Methods and Graphical Displays for Analyzing How the Association Between Two Qualitative Variables Differs Among Countries, Among Groups, Or Over Time: A Modified Regression-Type Approach », *Sociological Methodology*, vol. 28 (1), pp. 175-230.

Goux, Dominique et Éric Maurin (**2003**), « Who Marries Whom in France? », *in* Hans-Peter Blossfeld et Andreas Timm (dir.), *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, , pp. 57-78

Haller, Max (**1981**), « Marriage, Women, and Social Stratification: A Theoretical Critique », *The American Journal of Sociology*, vol. 86 (4), pp. 766-795.

Heinz, Ursula et Jan O. Jonsson (**2003**), « Who Marries Whom in Sweden? », *in Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, , pp. 235-266

Hou, Feng et John Myles (**2008**), « The Changing Role of Education in the Marriage Market: Assortative marriage in Canada and the United States since the 1970s », *Canadian Journal of Sociology*, vol. 33 (2), pp. 337-366.

Hout, Michael (**1982**), « The Association between Husbands' and Wives' Occupations in Two-Earner Families », *The American Journal of Sociology*, vol. 88 (2), pp. 397-409.

Janssen, Jacques P. G. (**2001**), *Do opposites attract divorce? Dimensions of mixed marriage and the risk of divorce in the Netherlands*, Purdue (États-Unis), Purdue University Press.

Johnson, Robert Alan (**1981**), *Religious Assortative Marriage in the United States*, New York, Academic Press.

Kalmijn, Matthijs (**1991**), « Status Homogamy in the United States », *The American Journal of Sociology*, vol. 97 (2), pp. 496-523.

Kalmijn, Matthijs (**1994**), « Assortative Mating by Cultural and Economic Occupational Status », *The American Journal of Sociology*, vol. 100 (2), pp. 422-452.

Kalmijn, Matthijs (**1998**), « Intermarriage and Homogamy: Causes, Patterns, Trends », *Annual Review of Sociology*, vol. 24, pp. 395-421.

Katrňák, Thomas (**2008**), « Educational Assortative Mating in the Czech Republic, Slovakia and Hungary between 1976 and 2003 », *Sociológia*, vol. 40 (3).

Kellerhals, Jean, Noëlle Languin, Jean-François Perrin et Geneviève Wirth (**1985**), « Statut social, projet familial et divorce : une analyse longitudinale des ruptures d'union dans une promotion de mariages », *Population*, vol. 40 (6), pp. 811-825.

Mare, Robert D. (**1991**), « Five decades of educational assortative mating », *American Sociological Review*, vol. 56 (1), pp. 15-32.

Mare, Robert D. (**2008**), « Educational Assortative Mating in Two Generations », *Department of Sociology, University Of California Los Angeles*.

Mare, Robert D. et Christine R. Schwartz (**2006**), « Educational Assortative Mating and the Family Background of the Next Generation », *Sociological Theory and Methods*, vol. 21 (2), pp. 253-278.

Qian, Zhenchao (**1998**), « Changes in Assortative Mating: The Impact of Age and Education, 1970-1990 », *Demography*, vol. 35 (3), pp. 279-292.

Raftery, Adrian E. (**1995**), « Bayesian Model Selection in Social Research », *Sociological methodology*, vol. 25, pp. 111-196.

Raymo, James M. et Miho Iwasawa (**2005**), « Marriage Market Mismatches in Japan: An Alternative View of the Relationship between Women's Education and Marriage », *American Sociological Review*, vol. 70 (5), pp. 801 -822.

De Saint Martin, Monique (**1985**), « Les stratégies matrimoniales dans l'aristocratie », *Actes de la recherche en sciences sociales*, vol. 59 (1), pp. 74-77.

Schwartz, Christine R. et Robert D. Mare (**2005**), « Trends in Educational Assortative Marriage from 1940 to 2003 », *Demography*, vol. 42 (4), pp. 621-646.

De Singly, François (**1982**), « Mariage, dot scolaire et position sociale », *Economie et statistique*, vol. 142 (1), pp. 7-20.

De Singly, François (**1987**), « Théorie critique de l'homogamie », *L'Année sociologique*, vol. 37, pp. 181-205.

Smits, Jeroen (**2003**), « Social Closure Among the Higher Educated: Trends in Educational Homogamy in 55 Countries », *Social Science Research*, vol. 32 (2), pp. 251-277.

Smits, Jeroen, Wout Ultee et Jan Lammers (**1998**), « Educational Homogamy in 65 Countries: An Explanation of Differences in Openness Using Country-Level Explanatory Variables », *American Sociological Review*, vol. 63 (2), pp. 264-285.

Ultee, Wout C. et Ruud Luijkx (**1990**), « Educational Heterogamy and Father-to-Son Occupational Mobility in 23 Industrial Nations: General Societal Openness or Compensatory Strategies of Reproduction? », *European Sociological Review*, vol. 6 (2), pp. 125-149.

Vallet, Louis-André (**1986**), « Activité professionnelle de la femme mariée et détermination de la position sociale de la famille. Un test empirique : la France entre 1962 et 1982 », *Revue française de sociologie*, vol. 27 (4), pp. 655-696.

Vallet, Louis-André (**2007**), « Sur l'origine, les bonnes raisons de l'usage, et la fécondité de l'odds ratio », *Courrier des statistiques*, n°121-122, pp. 59-65.

Vanderschelden, Mélanie (**2006a**), « Homogamie socioprofessionnelle et ressemblance en termes de niveau d'études : constat et évolution au fil des cohortes d'unions », *Economie et statistique*, vol. 398 (1), pp. 33-58.

Vanderschelden, Mélanie (**2006b**), « L'écart d'âge entre conjoints s'est réduit », *Insee première*, n°1073.

Verbakel, Ellen (**2008**), *The Partner as Resource or Restriction? Labour market careers of husbands and wives and the consequences for inequality between couples*, Nijmegen (Pays-Bas), Radboud University.

Verbakel, Ellen et Paul M. De Graaf (**2007**), « The association between spouses' occupational status: developments over the life course and over cohorts »,

Xie, Yu (**1992**), « The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables », *American Sociological Review*, vol. 57 (3), pp. 380-395.

Yamaguchi, Kazuo (**1987**), « Models for Comparing Mobility Tables: Toward Parsimony and Substance », *American Sociological Review*, vol. 52 (4), pp. 482-494.