



HAL
open science

Les rouages de l'amour et du hasard : homogamie et hypergamie dans la France et l'Europe contemporaines : dimensions socioéconomique et d'éducation, variations et mécanismes

Milan Bouchet-Valat

► **To cite this version:**

Milan Bouchet-Valat. Les rouages de l'amour et du hasard : homogamie et hypergamie dans la France et l'Europe contemporaines : dimensions socioéconomique et d'éducation, variations et mécanismes. Sociologie. Institut d'études politiques de paris - Sciences Po, 2015. Français. NNT : 2015IEPP0036 . tel-03499623v2

HAL Id: tel-03499623

<https://theses.hal.science/tel-03499623v2>

Submitted on 21 Dec 2021

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

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

Institut d'études politiques de Paris
ÉCOLE DOCTORALE DE SCIENCES PO
Programme doctoral de sociologie
Observatoire sociologique du changement

Doctorat en sociologie

Les rouages de l'amour et du hasard

*Homogamie et hypergamie
dans la France et l'Europe contemporaines :
dimensions socioéconomique et d'éducation, variations et mécanismes*

Milan Bouchet-Valat

*Thèse dirigée par Louis-André Vallet,
Directeur de recherche au CNRS*

Soutenue le 8 décembre 2015

Jury :

M. Francesco Billari, Professor, University of Oxford, Nuffield College

M. Philippe Coulangeon, Directeur de recherche, CNRS, OSC

M. Brendan Halpin, Senior Lecturer, University of Limerick (rapporteur)

M^{me} Éva Lelièvre, Directrice de recherche, Ined (rapporteure)

M. Louis-André Vallet, Directeur de recherche, CNRS, OSC (directeur de thèse)

SOMMAIRE

Remerciements.....	5
Introduction.....	11
Partie I : Dimensions et évolutions de l’homogamie en France (1969-2011).....	53
Chapitre 1 : De la structure de la population à celle des couples : le déclin de l’homogamie absolue de diplôme, de classe et d’origine sociales.....	55
Chapitre 2 : L’homogamie relative de diplôme, de classe et d’origine sociales et ses évolutions : ouverture d’ensemble, repli des élites.....	123
Chapitre 3 : La situation sur le marché du travail : effet égalisateur de l’activité féminine mais stabilité du lien entre statuts d’emploi.....	183
Partie II : Mécanismes de l’homogamie et de l’hypergamie en France.....	237
Chapitre 4 : La disparition de l’hypergamie relative de diplôme, de classe et d’origine sociales : une convergence du célibat entre les sexes (1969-2011).....	239
Chapitre 5 : Les mécanismes du célibat définitif des diplômées : la fin de la Sainte-Catherine (cohortes 1920-1970).....	291
Chapitre 6 : Les déterminants biographiques de l’homogamie et de l’hypergamie de diplôme au sein des premiers couples : l’âge et la scolarisation (cohortes 1920-1970).....	325
Partie III : Variations et déterminants de l’homogamie et de l’hypergamie en Europe.....	395
Chapitre 7 : La structure et le niveau de l’homogamie et de l’hypergamie socioéconomique et d’éducation : des différences marquées entre et à l’intérieur des pays.....	397
Chapitre 8 : Les facteurs de l’homogamie et de l’hypergamie socioéconomique et d’éducation : ouverture sociale et libéralisme culturel plus que niveau de développement.....	459
Conclusion.....	503
Annexes.....	515
Annexe A : Le codage de la classe Erikson-Goldthorpe-Portocarero (EGP) dans les enquêtes Emploi.....	517
Annexe B : Le coefficient d’association intrinsèque et l’ <i>odds ratio</i> absolu moyen (MAOR), deux mesures de l’association indépendantes des marges.....	525
Bibliographie.....	563
Liste des figures.....	595
Liste des tableaux.....	599
Liste des encadrés.....	601
Table des matières.....	603

REMERCIEMENTS

Je voudrais bien sûr tout d'abord remercier Louis-André Vallet, qui a su judicieusement m'orienter vers un sujet à propos duquel j'aurais cru que tout avait été dit, et qui a fait preuve d'une attention et d'une disponibilité sans failles tout au long de mon Master puis de ma thèse. Toutes les erreurs qui subsistent dans ce manuscrit ont été insérées à son insu. Merci aussi à Éva Lelièvre, qui m'a initialement accueilli au sein de l'unité Mobilité, logement et entourage de l'Ined à l'occasion d'un stage de Master afin d'exploiter les données de l'enquête Biographies et entourage, et qui n'a jamais renoncé à m'encourager à publier un article que je n'ai finalement terminé que trois ans – et presque une thèse – plus tard.

Cette thèse n'aurait pu être menée à bien sans le soutien des laboratoires qui m'ont accueilli pendant ces quatre années. J'ai pu rejoindre l'Observatoire sociologique du changement et l'École doctorale de Sciences Po grâce au soutien d'Alain Chenu, alors directeur de l'OSC. Je tiens aussi à remercier tous ses membres pour les remarques qu'ils ont pu me faire à l'occasion de présentations ou de discussions informelles.

Le soutien logistique, tant en termes de bureaux que de financement, m'a d'abord été offert par un contrat doctoral de trois ans au sein du Laboratoire de sociologie quantitative du CREST. Merci à Michel Gollac de m'avoir intégré au sein de cette équipe qu'il m'a d'emblée autorisé à tutoyer, et dont la cohésion compense largement la petite taille ; ainsi qu'à ses successeurs à la direction du LSQ, Dominique Goux puis Ivaylo Petev, pour avoir accepté la prolongation de mon affiliation et de l'accès à l'un des plus confortables bureaux de doctorants qu'il m'ait été donné de fréquenter. Merci aussi aux membres de ce laboratoire, en particulier à Fanny Bugeja, Danièle Guillemot, Yannick Lemel, Émilie Pauly, Juliette Stehlé, Sorana Toma et Sander Wagner.

En plus d'un accueil à temps partiel tout au long de ma thèse, l'Ined m'a assuré un financement de quatrième année qui m'a permis de terminer ce travail dans des conditions idéales. J'ai eu la chance de présenter plusieurs fois mes travaux aux membres de l'unité Mobilité, logement et entourage. Au service des méthodes statistiques, Élisabeth Morand m'a apporté son éclairage concernant la régression PLS, et Bénédicte Garnier m'a encouragé et aidé à travailler sur le projet R.TeMiS plutôt que sur cette thèse : qu'elles en soient également remerciées.

Les meilleures conditions matérielles et scientifiques ne seraient rien sans le soutien et la bonne entente quotidienne entre doctorants. À cet égard, la mise en place d'un séminaire doctoral autogéré (SDAG) au sein du LSQ – et sa variante BeT –, a simultanément contribué à l'ambiance et à la qualité littéraire et scientifique de cette thèse. Merci donc pour leurs relectures et débats interminables, « leur rigueur et leur rigolade » à (par ordre approximatif d'apparition à l'écran) Marie-Paule Couto, Joanie Cayouette-Remblière, Noémie Le Donné, Yoann Demoli, Mathieu Ichou, Julien Gros, Camille Martin, Rémi Sinthon, Nicolas Deffontaines, Marion Plault, Lucas Tranchant, Pauline Vallot, Marine Haddad et Haley Mcavay.

Merci aussi aux doctorants les plus assidus de l'OSC qui ne m'en ont pas (trop) voulu de ne pas accorder au laboratoire une fréquentation régulière, dont notamment Annabelle Allouch, Lorraine Bozouls, Pauline Clech, Margot Delon, Alice Olivier et Agathe Voisin. Parmi eux, Marie Bergström a accepté avec enthousiasme une collaboration – toujours repoussée – sur les sujets qui nous sont communs.

Les doctorants de l'Ined ne sont pas non plus en reste. Puisque leur nombre m'empêche de les citer tous, qu'il me soit permis de ne mentionner que ceux avec lesquels j'ai le plus fréquemment partagé un bureau, un Master, une unité, un étage, des idées, des repas ou simplement des pauses : Élixa Abassi, Julie Chapon, Pierre-Antoine Chauvin, Mona Claro, Noémie De Andrade, Manon Garrouste, Anne Gosselin, Amélie Grysole, Pauline Hervois Sandrine Juin, Morgan Kitzmann, Vincent Lignon, Celio Sierra-Paycha, Valeria Solesin, Valentine Trépied, Thomas Venet, Paola Villar.

Plusieurs chercheurs ont accepté de discuter mes présentations à l'occasion du séminaire interne du LSQ et de la journée doctorale de l'Ined : Frédéric Lebaron, François Héran, Michel Bozon puis Laurent Toulemon. Monique Meron et Michel Amar m'ont permis

de participer au projet ESeG, qui m'a ouvert l'accès à la classification et aux données utilisées dans la troisième partie de cette thèse. Enfin, le réseau Quetelet a fourni nombre des bases de données utilisées ici, et de nombreux anonymes ont contribué à la construction et à la collecte de ces enquêtes.

Merci à Aurélie qui, en tant que colocataire durant ma première année de thèse, a sans doute d'une manière ou d'une autre contribué à la définition de sa problématique.

Merci à mes parents et grands-parents, dont le soutien ne fait pas mentir les théories sociologiques concernant le lien entre origine sociale et niveau d'études.

Merci enfin à Maeva qui a prouvé – rayer la mention inutile – la valeur de l'hypergamie féminine en assurant l'intégralité des tâches domestiques pendant la préparation de cette thèse ; ou bien celle de l'homogamie en corrigeant assidûment l'ensemble de ses chapitres.

Vous trouverez chez moi la belle Hélène qu'on ne se lasse jamais de voir. Lorsqu'il lut cette phrase, Pierre Bezoukhof comprit, pour la première fois, qu'il existait, entre Hélène et lui, une sorte de liaison déjà admise par tout le monde. Le fait l'effraya, car il lui imposait des obligations qu'il ne pouvait remplir, mais il lui plut d'y penser comme à une plaisante éventualité. [...] Anna Pavlovna rendit sa liberté à Pierre, qui, très sincèrement, partagea son avis quant à la tenue et au ton d'Hélène. Lorsqu'il pensait à cette princesse, c'était toujours sa beauté qu'il évoquait et son art extraordinaire de prendre en société une attitude silencieuse, calme et digne. [...] Il devenait impossible de voir Hélène autrement après cette découverte, de même que nous ne pouvons être trompés, une seconde fois, par une supercherie. [...] Et, incontinent, Bezoukhof sentit que non seulement Hélène pouvait, mais qu'elle devait être sa femme et qu'il ne pouvait en être autrement. Comment et quand cela se ferait-il ? Pierre l'ignorait. Serait-ce un événement heureux ? Il prévoyait confusément le contraire. Mais sans aucun doute, l'événement aurait lieu !

Léon Tolstoï (1869), *Guerre et paix*, I, III,
traduction Boris de Schloezer.

INTRODUCTION

La mise au jour des déterminants sociaux du choix du conjoint peut être considérée, à l'instar de l'étude du suicide d'Émile Durkheim, comme l'une des démonstrations classiques de la force de l'approche sociologique. En révélant le caractère socialement situé d'une décision profondément intime que les sociétés occidentales modernes regardent comme le lieu par excellence où s'expriment l'individualité et le hasard, la sociologie souligne l'influence essentielle de la stratification sociale sur les comportements individuels. En France, les travaux désormais classiques d'Alain Girard (1964) puis de Michel Bozon et François Héran (2006) ont offert une vision saisissante des mécanismes par lesquels les conjoints se « découvrent » et s'apprécient : au-delà même des lieux de rencontre, les critères de jugement des partenaires potentiels – et en particulier ceux portant sur l'apparence physique – diffèrent entre groupes sociaux. Le scandale que vient rappeler la sociologie à propos du choix du conjoint tient en un mot : l'amour est loin d'être aveugle aux distinctions sociales.

La centralité de la question du choix du conjoint dans la stratification sociale a elle aussi été reconnue de longue date par la discipline. Max Weber ([1922] 1971, partie 1, ch. IV et partie 2, ch. V, et [1922] 1968, partie 2, ch. IX, sec. 6) a ainsi mis en avant son importance dans la constitution et le maintien de l'identité des *groupes de statut*. Dans ses termes, les alliances conjugales (*connubium*) et la commensalité (*convivium*, littéralement le fait partager un repas, mais plus largement l'amitié) expriment le fait que les individus ou les familles concernées se considèrent comme des égales, et confirme leur similarité de styles de vie.

À la suite de Weber et d'autres auteurs classiques, l'homogamie – soit la tendance à choisir des conjoints socialement proches de soi – a fréquemment été considérée comme un point d'entrée privilégié pour l'étude de la stratification sociale dans la littérature

internationale (voir Kalmijn, 1998 ; Blossfeld, 2009 ; Schwartz, 2013 pour des revues). En effet, elle est à la fois symptôme, cause et conséquence de l'existence d'une stratification sociale et de groupes sociaux dotés d'une cohérence, sous le double aspect de l'en-soi et du pour-soi : en-soi, puisque les conditions de vie objectives déterminent les probabilités de rencontre et de partage de goûts communs entre potentiels conjoints ; pour-soi, puisque le statut social détermine la possibilité même de se mettre en couple, et puisque l'intensité des liens entre individus partageant une même condition objective peut être un facteur majeur de cristallisation d'une identité de groupe.

En premier lieu, l'homogamie peut être considérée comme un *symptôme* de la structure sociale, en termes à la fois objectifs et subjectifs. En effet, pour que deux individus puissent former un couple, ils doivent nécessairement avoir eu l'occasion de se rencontrer, et partager une perception commune du lieu, du moment et de la manière de le faire (Bozon et Héran, 2006, chap. 1 et 2 ; Kalmijn et Flap, 2001) ; plus généralement, ils doivent présenter une certaine proximité spatiale et culturelle. Au-delà, de telles rencontres ne peuvent donner lieu à la formation d'un couple sans le sentiment de la part des deux conjoints potentiels que leurs statuts sociaux sont assez proches pour rendre la formation d'un couple concevable. En étendant la lecture wébérienne, on peut interpréter la propension d'individus appartenant à des groupes socialement éloignés à former des couples ensemble comme un indice du degré d'ouverture d'une société, et, en négatif, de l'importance qu'y prend le statut social. À l'échelle d'un groupe particulier, une endogamie élevée est un facteur de maintien d'une identité forte, pouvant en particulier signaler une congruence entre classe et groupe de statut (« identité temporelle » d'une classe sociale suivant Chauvel, 2001) ; à l'inverse, une baisse de l'endogamie peut à la fois signaler et entraîner un déclin de la conscience de classe, avec des effets potentiellement négatifs en termes d'intégration sociale.

Ainsi, suivant un mécanisme en partie circulaire, l'homogamie est en retour un *déterminant* majeur de la structure sociale, confirmant ou infirmant l'échelle de statut en vigueur, et aggravant ou atténuant les inégalités, *via* la mise en commun du capital économique, culturel et social au sein d'un ménage. Les inégalités entre familles peuvent à leur tour avoir de grandes conséquences sur celles qui s'observeront au sein de la génération suivante (Mare et Schwartz, 2006 ; Schwartz, 2013). Enfin, des inégalités importantes auront tendance à renforcer les différences de styles de vie entre groupes sociaux, ainsi qu'à rendre

économiquement et culturellement plus coûteuse une alliance hors du groupe, faisant de l'homogamie une *conséquence* potentielle d'une stratification fortement marquée par l'appartenance de classe.

Le choix du conjoint, parce qu'il implique – du moins dans les couples hétérosexuels – un homme et une femme, met aussi en jeu les rôles de genre, et permet leur confirmation ou leur remise en cause. En effet, à côté d'une tendance à former des couples entre individus socialement proches, on relève une asymétrie entre les positions des hommes et des femmes, qui prend la forme de l'hypergamie féminine – soit la supériorité du statut social de l'homme au sein du couple – ou de manière équivalent de l'hypogamie masculine – le même phénomène pris du point de vue des hommes (de Singly, 1987a ; Guichard-Claudic, Testenoire et Trancart, 2009)¹. Comme l'homogamie exprime et renforce l'ordre social, l'hypergamie féminine signale et favorise le maintien de la domination masculine en mettant la femme en infériorité dans nombre de situations sociales, et en particulier en ce qui concerne le niveau d'éducation et la carrière professionnelle. Cet ordre de genre implique en outre des conséquences notables en termes de célibat, les individus ayant le moins de chances d'arriver à respecter ce schéma asymétrique – soit les hommes les moins diplômés d'un côté, et les femmes les plus diplômées de l'autre – éprouvant des difficultés à trouver un partenaire acceptable pour former un couple (de Singly, 1987b, chap. 8).

Les deux logiques de choix du conjoint – de statut et de genre, de symétrie et d'asymétrie – gagnent à être traitées ensemble. En effet, comme nous tâcherons de le montrer, homogamie et hypergamie varient souvent dans le même sens et ont de nombreux déterminants en commun. En adoptant et en développant, concernant la France au cours des dernières décennies et l'Europe actuelle, une approche bien établie au sein de la littérature internationale, mais moins connue dans la littérature française, cette thèse cherchera à démontrer l'intérêt de l'étude du choix du conjoint à l'aide de grandes enquêtes quantitatives pour appréhender les transformations au cours du temps et les variations dans l'espace des sociétés à la fois en termes de stratification sociale et de rapports entre les sexes. Seule une analyse quantitative – dont les limites sont par ailleurs évidentes dans

1 Suivant la convention retenue dans la littérature, nous parlerons généralement d'hypergamie féminine, ou plus simplement d'hypergamie, sans impliquer que ce phénomène doive être attribué aux femmes plus qu'aux hommes.

d'autres domaines – permet en effet de mesurer de manière précise l'homogamie et l'hypergamie, et de saisir leurs déterminants au niveau macrosociologique.

Nos analyses porteront d'une part sur les évolutions du choix du conjoint en France au cours des quarante dernières années ainsi qu'au fil des cohortes nées entre 1920 et 1970, et d'autre part sur l'Europe des années récentes. Le cas de la France présente plusieurs particularités qui en font un terrain intéressant pour mettre à l'épreuve plusieurs des théories avancées par la littérature, que nous développerons plus bas. Premièrement, ce pays a été assez tardivement agricole, ce qui permet d'observer à partir de données d'enquête qui ne sont pas disponibles pour les périodes plus anciennes les derniers signes des effets de l'industrialisation et de l'exode rural sur le choix du conjoint. Deuxièmement, il a fait l'expérience d'une massification scolaire particulièrement marquée sur la période étudiée, qui a fortement affecté la structure des relations entre groupes et le choix du conjoint. Troisièmement, il a connu une forte hausse du niveau d'études et du taux d'activité des femmes, mais aussi une augmentation de la continuité des carrières féminines, qui ont eu de nets effets sur l'hypergamie, mais aussi l'homogamie. En outre, en termes pratiques, la société française a été couverte sur le long terme par des enquêtes quantitatives de qualité et aux effectifs élevés. Pourtant, s'il existe en France une riche tradition d'étude concernant les modalités de la rencontre du conjoint, les travaux s'attachant à mesurer les évolutions de l'homogamie et de l'hypergamie sont peu nombreux, et les connaissances parcellaires. Cette thèse vise à contribuer à combler ce manque ; dans le cas de l'homogamie comme de l'hypergamie, nos analyses viennent remettre en cause plusieurs évidences apparentes généralement tenues pour acquises par la sociologie française.

L'étude des évolutions dans un pays donné ne permet pas véritablement de démêler l'influence des multiples facteurs qui évoluent fréquemment dans le même sens au cours du temps. C'est pourquoi nous porterons ensuite notre attention sur les variations du choix du conjoint entre les régions composant les pays d'Europe. Cette étude permettra de situer la France dans le contexte d'un ensemble de pays relativement proches, mais suffisamment dissemblables pour permettre une analyse des variations de l'homogamie et de l'hypergamie ainsi que de leurs déterminants. L'apport de cette étude ne tiendra pas à l'extension de son champ géographique ou temporel, des études bien plus larges ayant déjà été réalisées. Il proviendra au contraire de sa restriction, qui assure une grande comparabilité des pays

étudiés, et permet de disposer de données de qualité rendant possible une analyse détaillée des facteurs du choix du conjoint. Les résultats permettront de mettre à l'épreuve les théories existantes concernant les déterminants du choix du conjoint aux stades avancés du développement.

Dans cette introduction, nous commençons par rappeler le consensus qui prévaut peu ou prou au sein de la sociologie française concernant les deux aspects du choix du conjoint, et ébauchons des critiques à lui adresser. Nous présentons ensuite successivement les évolutions historiques de long terme qui déterminent le choix du conjoint : d'une part, le déclin du contrôle social sur le choix du conjoint et l'importance croissante accordée au sentiment amoureux ; d'autre part, et en lien avec la précédente tendance, le processus d'individualisation identifié dès les origines de la sociologie comme constitutif de la modernité. Puis nous précisons la forme que prennent ces processus historiques dans le contexte de notre étude sur des sociétés européennes contemporaines, à travers l'évolution récente des modèles familiaux et de la stratification sociale. Enfin, nous présentons le cadre d'analyse de l'homogamie et de l'hypergamie généralement retenu dans la littérature, avant de le décliner en trois ensembles de mécanismes que nous mobiliserons tout au long de cette thèse. Nous concluons par une présentation du plan de la thèse.

1 Des évidences bien fragiles

Le tour de force des études françaises classiques sur le choix du conjoint tient au constat d'une persistance de l'homogamie sociale en dépit des transformations majeures qu'a connues la société, et tout particulièrement la famille. A. Girard (1964) frappait – ou rassurait – ainsi son époque en montrant le décalage entre le sentiment de fort accroissement de la liberté dont jouissaient les jeunes en matière conjugale depuis la fin de la Seconde Guerre mondiale, et le maintien de l'ordre social qui résultait malgré tout de l'agrégation de ces choix théoriquement libres et potentiellement aléatoires. M. Bozon et F. Héran (2006) n'étonnaient pas moins, en confirmant ce constat au milieu des années 1980 dans un contexte d'individualisation des modèles familiaux. Ils accentuaient même le saisissement en démontant les mécanismes de la persistance de l'hypergamie féminine après plusieurs décennies de remise en cause des rôles de genre. Récemment, M. Bozon et Wilfried

Rault (2012) ont étendu ce résultat au choix du premier partenaire sexuel, alors que Marie Bergström (2014) a identifié les canaux par lesquels la sélection sociale du conjoint s'opère jusque sur les sites de rencontres en ligne.

1.1 Un maintien de l'homogamie ?

Cette tradition de travaux a donné lieu au constat maintes fois repris de la persistance de l'homogamie et de l'hypergamie dans la société française contemporaine. De ce résultat majeur, on a sans doute un peu vite conclu que la détermination sociale du choix du conjoint n'avait pas *varié* au cours du temps. Ainsi, on trouve déjà sous la plume de Pierre Bourdieu une théorie du transfert d'un mode de sélection sociale à un autre qui permet le maintien, voire le renforcement, de l'homogamie d'éducation :

Avec les progrès de l'accès des filles à l'enseignement supérieur, les mécanismes d'auto-orientation (« vocation ») et de sélection qui produisent des groupes scolaires (faculté ou école, discipline, etc.) socialement très homogènes ont tendu à assurer l'endogamie de classe (ou de fraction) au moins aussi efficacement, mais selon un tout autre mode, que l'interventionnisme des familles et en particulier leurs efforts pour organiser les occasions directement contrôlées de rencontre (bals, surprise parties, rallyes, etc.). Cet effet inattendu de la scolarisation n'a sans doute pas peu contribué à encourager les familles à abandonner leur politique dirigiste (en tout cas fort difficile à imposer) au profit du laisser-faire en même temps que se redéfinissait complètement le système des critères qui déterminaient la valeur des filles sur le marché matrimonial, qu'il s'agisse du capital économique (dot) ou du capital d'honorabilité (virginité, maintien, etc.).

(Bourdieu, 1974, p. 35)

Point caractéristique, cette affirmation se fonde pour partie (d'après une note de bas de page) sur des travaux empiriques portant sur les États-Unis. Or, quarante années plus tard, c'est la même analogie géographique qui est parfois reprise pour avancer que l'homogamie d'éducation se renforce en France. La force du consensus est telle qu'Henri Mendras, dans *La Seconde Révolution française* – livre pourtant considéré comme la défense la plus affirmée de la thèse de la moyennisation de la société – voyait l'apparition d'isolats professionnels du fait de l'augmentation de l'homogamie parmi les plus diplômés, et soulignait la force de l'homogamie aux deux extrêmes de l'échelle sociale (Mendras, 1988, p. 52).

Les manuels de référence sur la sociologie du couple abondent dans ce sens. On lit par exemple que « *sans explicitation et sans calcul, les familles contemporaines parviennent ainsi à obtenir d'aussi bonnes alliances du point de vue social que les familles traditionnelles.* » (de Singly, [1993] 2014, p. 38) ou que « *la fin des mariages arrangés n'a pas diminué [...] l'homogamie* » (de Singly, 1988, p. 131, voir aussi 1996a). S'il est parfois admis que l'homogamie en termes professionnels ait pu diminuer, c'est pour aussitôt affirmer qu'« *il est vraisemblable qu'il s'agisse d'un simple glissement de critères, le métier devenant moins discriminant que le niveau de revenu, la position sociale, les affinités culturelles* » (Kaufmann, [1993] 2014, p. 6). La force persistante de l'homogamie frappe jusqu'à des auteurs qui avancent par ailleurs tous les arguments qui expliqueraient son affaiblissement (Kellerhals *et al.*, 1982, p. 53).

Ce paradoxe est pourtant relativement simple à résoudre : du constat de l'existence d'un phénomène, on ne saurait déduire sa stabilité temporelle ; ce serait singulièrement manquer d'imagination sociologique que de s'interdire de penser que l'homogamie ait pu être *encore plus forte* par le passé. Comme nous le verrons, certaines preuves empiriques indiquant une tendance à l'affaiblissement de l'homogamie en France ont été relevées depuis une trentaine d'années, même si des interrogations demeurent. Cette thèse apportera une confirmation de ce mouvement, qui nous semble paradoxalement plus congruent avec les théories développées par les auteurs déjà cités que leurs affirmations concernant son maintien à l'identique.

1.2 Une hypergamie inébranlable ?

Une autre dimension du choix du conjoint fait l'objet d'un consensus discutable : l'hypergamie et la dimension genrée du célibat. Un apport important des études de ce champ a été de mettre en évidence que la proximité sociale des conjoints se conjugait avec une asymétrie marquée entre hommes et femmes que la notion d'homogamie amène à occulter (de Singly, 1987a).

Cependant, dans un contexte de remise en cause des rôles de genre et de transformations rapides des modèles familiaux, faire reposer encore aujourd'hui les théories du choix du conjoint sur des données datant des années 1980 apparaît pour le moins risqué.

Il peut sembler étonnant que les présentations des manières genrées de valoriser ses atouts sur le marché conjugal fassent encore presque exclusivement référence à l'exemple des annonces matrimoniales parues dans le *Chasseur français* à cette période (de Singly, 1984), alors que ce mode de rencontre est extrêmement minoritaire et ne saurait être considéré comme représentatif de la majorité des mises en couple. Dans les années récentes, les sites de rencontre en ligne ont profondément renouvelé ce mode de recherche d'un conjoint (Bergström, 2011, 2014).

De fait, nous montrerons que la structure genrée des couples et les normes de genre régissant le choix du conjoint ont nettement évolué au cours des dernières décennies : les françaises sont désormais plus diplômées que leurs conjoints ; et, si elles occupent toujours des emplois inférieurs à ces derniers, le choix du conjoint n'y est plus pour rien. Enfin, les différences de célibat se sont en grande partie résorbées, et les diplômées ne souffrent plus d'aucun handicap sur le marché conjugal.

2 L'invention de l'amour ?

Le scepticisme concernant un potentiel affaiblissement de l'homogamie apparaît cependant en partie comme une saine réaction de la sociologie à un discours trop fréquemment caricatural concernant le passage du mariage arrangé à un choix absolument libre du conjoint, avec le cœur comme seul conseiller. En effet, si nous acceptons la position selon laquelle la modernité est associée à une moindre détermination sociale du choix du conjoint, ce n'est pas tant du fait d'une supposée invention de l'amour que d'une transformation de la place qui est accordée à ce sentiment, du développement de l'individualisme et de l'affaiblissement du contrôle social.

2.1 Une « révolution sentimentale » surestimée

Ainsi, William Goode (1959, [1964] 1982, chap. 4) montre que le sentiment amoureux est présent dans toutes les sociétés, qui diffèrent avant tout par la place qui lui est attribuée ainsi que par le degré et les modalités du contrôle social qui lui est appliqué. Le choix du conjoint est généralement plus libre parmi les classes inférieures – par exemple, parmi les paysans de l'Europe préindustrielle –, mais plus contrôlé parmi les classes supérieures, qui

disposent des ressources nécessaires et chez lesquelles les enjeux liés au mariage sont fondamentaux. Ce contrôle peut varier en intensité, allant d'une ségrégation complète des jeunes (et notamment des jeunes femmes)², à de simples remarques, en passant par une action plus ou moins directe sur les fréquentations ou une surveillance lors des sorties (chaperonnage). Les sociétés modernes ne sont pas de nature radicalement nouvelle de ce point de vue : le contrôle social du choix du conjoint y a seulement perdu en intensité, les pressions s'exerçant plutôt sous la forme d'incitations de la part des parents ou du groupe de pairs.

Même lorsque l'amour n'est pas considéré comme le fondement du couple, son expression n'est pas nécessairement réprimée. Paul Veyne (1978, p. 41) a par exemple décrit en détail la place de la passion amoureuse dans l'aristocratie du Haut-Empire romain : contrairement à la conception moderne, celle-ci n'est concevable que dans le cadre de relations extra-conjugales, largement tolérées. C'est bien la valorisation de l'amour comme seul fondement acceptable du couple qui différencie les sociétés modernes de nombre d'autres sociétés connues. W. Goode désigne sous le nom de « *complexe de l'amour romantique* » la forte institutionnalisation de l'amour dans le cadre du mariage observée dans les sociétés occidentales contemporaines.

À l'inverse de cette thèse, certains historiens, dans des ouvrages encore largement cités (Shorter, [1975] 1977 ; Stone, 1977), ont affirmé que le sentiment amoureux était caractéristique des sociétés modernes, et que son apparition avait eu lieu en Europe à partir du XVII^e siècle. Ces analyses se sont attelées à la tâche délicate de dresser une histoire des sentiments couvrant toutes les catégories de la population, en dépit des difficultés liées à la rareté des archives traitant de questions intimes, en particulier pour les périodes les plus anciennes et les couches non lettrées – la majorité – de la société. Si l'objectif est louable, les conclusions de ces travaux nous semblent peu acceptables. Dès leur parution, de vives critiques leur ont été adressées. Ainsi, quatre revues du premier ouvrage (Matossian, 1976 ; Branca, 1977 ; Farber, 1977 ; Gordon, 1977) ont été particulièrement sévères à l'égard de la

2 Cette stratégie de ségrégation des sexes, reposant à la fois sur une séparation physique et sur l'éducation, est notamment décrite par Pierre Bourdieu (1962, p. 56) chez les paysans du Béarn des années 1960, chez lesquels le rôle des entremetteurs est central et les stratégies matrimoniales très explicites. Contre une lecture linéaire des évolutions historiques, J.-L. Flandrin (1975, p. 107) considère cette pratique comme relativement récente.

thèse d'une « *révolution sentimentale* » et de sa démonstration empirique. Jack Goody (2001, p. 213, voir aussi [1983] 1985, chap. I) a relevé l'« *ethnocentrisme marqué* » de ces travaux.

De fait, précisant la thèse de W. Goode et d'autres auteurs, on dispose de nombreuses illustrations de l'importance du sentiment amoureux dans l'Europe pré-moderne – que ce soit dans le cadre du mariage, en dehors de lui, voire contre lui. Critiquant encore une fois ces travaux, Didier Lett (2000, chap. 10) fournit ainsi des exemples d'amour conjugal en France au Moyen-Âge. Ferdinand Mount (1982, chap. 5), étudiant le mariage en Angleterre entre les XIV^e et XVI^e siècles, observe à travers des lettres et des jugements de tribunaux que, même au sein des strates privilégiées de la société, pour lesquelles les intérêts matériels dépendent largement des stratégies matrimoniales, l'amour entre futurs époux était apprécié lorsqu'il n'allait pas à l'encontre des premières. Il était aussi relativement fréquent que les enfants refusent les conjoints que leur destinaient leurs parents pour des raisons sentimentales. Alan Macfarlane (1986, part. III) a lui aussi observé dans l'Angleterre du XVII^e siècle, parmi le peuple autant que parmi les classes intermédiaires (comme les marchands), que les jeunes avaient l'initiative du mariage dans la plupart des cas, l'entourage ne faisant qu'encadrer le choix du conjoint. Dès le XVI^e siècle, des auteurs considèrent le consentement des futurs époux comme essentiel, et encouragent les enfants à résister aux pressions exercées par les parents.

L'importance accordée au consentement des époux est loin d'être un phénomène récent. Dès le XII^e siècle, l'Église s'est ainsi opposée aux familles et à l'État en soutenant que le consentement des époux était la seule base légitime du mariage (Goody, [1983] 1985, p. 217 ; Daumas, 2004). Elle reconnaissait même, tout en les condamnant, les mariages prononcés par serment des deux époux hors de toute cérémonie religieuse ou publique et sans consentement parental. Jean-Louis Flandrin (1975, chap. « Évolutions ») considère que l'insistance de l'État (particulièrement en France) à lutter contre les mariages sans consentement parental est le signe que ces pratiques étaient assez fréquentes pour constituer un problème, notamment pour la noblesse.

2.2 Des évolutions réelles

Il ne s'agit pour autant pas de nier que des évolutions majeures ont pris place dans les processus de choix du conjoint. J.-L. Flandrin (1975) affirme ainsi, en se fondant sur l'évolution des conceptions pré-nuptiales, que le contrôle des jeunes s'est relâché en France à partir du XVII^e siècle. Cette évolution est plus marquée dans les villes ainsi que dans les villages où l'industrie textile s'est fortement développée, que dans le reste du monde rural. L'existence de charivaris à l'encontre des filles qui se mariaient en dehors du village, organisés par les garçons qui voyaient leur échapper une épouse potentielle, suffit sans doute, par son contraste avec les pratiques contemporaines, à rappeler que les évolutions en la matière ont été notables.

Cette évolution est aussi relevée par Wally Secombe (1993, p. 18) parmi la classe ouvrière anglaise au cours de la Révolution industrielle. Pour l'auteur, le développement du capitalisme brise la continuité des générations, qui était auparavant assurée par la transmission des moyens de production au sein d'une famille qui était aussi une unité de production. Avec le développement du salariat, les jeunes échappent à l'influence de leurs parents et de leur communauté, à qui ils ne doivent plus leur emploi, ni leur logement – auquel le premier leur permet d'accéder par eux-mêmes. Détaché des contraintes matérielles et du contrôle social, le choix du conjoint devient donc relativement libre. Pour autant, les liens entre parents et enfants ne disparaissent pas avec la tendance nouvelle à la néolocalité ; ils affectent toujours le choix du conjoint, mais de manière indirecte.

Plus près de nous, les travaux de P. Bourdieu ou d'A. Girard donnent des illustrations de ces tendances dans la France du XX^e siècle, dessinant des pratiques qui peuvent nous paraître parfois plus étrangères que certaines lettres d'amour du XVII^e siècle reprises par les auteurs déjà cités. Ainsi, dans son étude des paysans du Béarn au début des années 1960, P. Bourdieu (1962) nous décrit un monde dans lequel l'héritage était jusqu'alors la considération principale présidant au choix du conjoint, et dans lequel les cadets – exclus de l'héritage par la règle de primogéniture – n'avaient d'autre choix que de s'exiler ou de se résigner au célibat. Le contrôle social sur la mise en couple y était très fort, du fait de la centralité des lignages et de la propriété foncière qui leur était associée, ainsi que de l'interconnaissance. C'est à la fin de ce monde qu'assiste le sociologue, sous la forme

notamment de l'augmentation du célibat des aînés, qui, tout héritiers qu'ils soient, font désormais face à la concurrence des citadins, porteurs d'un autre modèle conjugal.

La forte imbrication de la recherche du conjoint dans les réseaux locaux d'interconnaissance est aussi observée par A. Girard (1964, p. 198) dans la France de l'après-guerre. Si le choix du conjoint n'est pas soumis à l'accord parental, il n'en est pas pour autant indépendant de la génération précédente : « *Pour se trouver, il leur [les conjoints] faut dans une large mesure l'aide de la société, et comme la complicité des adultes.* » L'autonomisation croissante de la jeunesse (Galland, [1991] 2011) est venue progressivement défaire ce modèle, modifiant du même coup les chances de rencontre entre groupes sociaux. De fait, l'auteur observe (Girard, 1964, chap. III.1.7) que les générations récentes, qui sont aussi les plus jeunes, déclarent accorder moins d'importance à la similitude des milieux d'origine que leurs aînés. C'est aussi ce que relève un peu plus tard Louis Roussel (1975, sect. II.1.4).

2.3 Les nouvelles formes du contrôle social

Après avoir relativisé le caractère contraint du choix du conjoint dans l'Europe pré-industrielle, il n'est pas inutile de souligner, dans l'autre sens, la persistance de certaines formes de contrôle social dans les sociétés européennes contemporaines. L'influence parentale est moins directe, mais elle s'impose toujours en partie aux jeunes. Ceux-ci présentent ainsi rapidement leurs partenaires à leurs parents, et en attendent une forme de « caution » (Lemarchant, 2001). Bien que l'enjeu en soit moindre qu'auparavant, cette présentation garde un caractère ritualisé, source d'une certaine anxiété pour les deux conjoints. Les différences de milieu social sont parfois directement à l'origine de tensions entre le couple et les beaux-parents ; néanmoins, en cas d'opposition franche avec les parents, les normes sociales concernant la liberté du choix du conjoint et l'autonomie du couple peuvent paradoxalement renforcer sa cohésion face à l'extérieur.

Les intéressés reconnaissent volontiers l'importance de l'opinion de leurs parents concernant leur conjoint (Le Pape, 2006, p. 18). Valable nous semble-t-il pour tous les groupes sociaux, ce point a été particulièrement souligné en ce qui concerne la norme d'« homogamie socio-ethnique » au sein des minorités ethniques (Santelli et Collet, 2011 ;

voir aussi Clair, 2008). Ces deux auteurs parlent ainsi d'« endogamie élective » pour décrire l'intériorisation des attentes des parents concernant le mariage avec un conjoint appartenant à la même communauté d'origine, sans pour autant renoncer à fonder son choix sur le sentiment amoureux. Les réticences à aller à l'encontre des attentes parentales sont même à l'origine d'un célibat relativement fréquent parmi les jeunes femmes d'origine maghrébine – en particulier les plus diplômées –, qui peinent à concilier cette exigence avec un refus de se plier entièrement aux rôles de genre en vigueur dans ce groupe (Flanquart, 1999).

Dans un autre registre, le cas de l'aristocratie et de la haute bourgeoisie est lui aussi éclairant (Wagner, 2008). Si les mariages arrangés ont bel et bien disparu, l'influence parentale n'a pas pour autant complètement cédé dans ces milieux qui ont le plus à perdre à une mésalliance. Ainsi, la pratique du rallye (Merllié et Cousquer, 1980 ; Pinçon et Pinçon-Charlot, 1989 ; Wagner, 2008), qui vient en complément d'une ségrégation résidentielle déjà marquée, est apparue dans les années 1950. Loin de représenter un vestige de pratiques d'un autre âge, elle ne vise pas tant à contrôler directement les conjoints potentiels des jeunes, en restreignant leurs contacts à un réseau sélectionné par avance, qu'à leur enseigner les manières du milieu et à construire des relations qui serviront plus tard à trouver son conjoint. Grâce à ce dispositif de socialisation, l'intervention directe de l'entourage est rarement nécessaire – même si la menace d'une exclusion du groupe est présente dans les esprits.

Ces deux exemples très différents illustrent l'évolution du choix du conjoint qui affecte tous les groupes sociaux. L'encadrement social n'a pas totalement disparu ; il est néanmoins devenu plus discret, et au total moins prégnant. Il nous paraît difficile de soutenir que les conjoints ne sont pas plus libres de se choisir dans la France actuelle qu'il y a un siècle, ou même cinquante ans.

2.4 Les concepts de choix du conjoint et de marché conjugal

Ainsi, nous nous inscrivons en rupture avec les travaux faisant l'hypothèse de l'existence d'un calcul stratégique sur le marché conjugal. Si P. Bourdieu (1972) pouvait sans risque parler de « stratégies matrimoniales » à propos du choix du conjoint dans la société

paysanne du Béarn, dans laquelle ce calcul était omniprésent, et si cet usage est sans doute encore aujourd'hui valide parmi les classes possédantes (Wagner, 2008), son application généralisée au choix du conjoint au sein des sociétés occidentales contemporaines nous paraît abusive, niant l'importance des évolutions historiques ayant affecté les modèles conjugaux au cours du XX^e siècle. Ce serait faire bien peu de cas du sens que donnent les individus à leurs actions que de persister à confondre sous un même concept les stratégies de reproduction portant sur les investissements économiques ou l'orientation scolaire des enfants et le choix du conjoint (Bourdieu, 1974). Même si sa définition bourdieusienne stricte est bien plus subtile³, la notion de stratégie nous semble trop facilement renvoyer à un calcul conscient, visant une définition trop restreinte de l'intérêt, pour être véritablement utile à notre analyse. Ce risque de glissement sémantique apparaît d'autant plus fort que, comme la plupart des travaux quantitatifs, nous n'avons accès qu'au *résultat* des actions individuelles, et pas à leurs motifs conscients, ce qui amène trop fréquemment à revenir à l'interprétation la plus facile en termes d'intérêt et de calcul.

Il ne s'agit pas pour autant de nier que l'intérêt puisse être présent, consciemment ou non, sous les apparences du sentiment, et qu'il s'infilte jusque dans les critères d'appréciation apparemment les plus anodins. Mais il ne correspond qu'exceptionnellement à une tentative de maximisation du statut social ou des revenus du conjoint que certains travaux placent derrière un tel vocabulaire. Dans un contexte de valorisation croissante du lien conjugal et de l'individualité, les gains affectifs ou identitaires, qui découlent notamment d'une communauté de vues au sein du couple, peuvent s'avérer bien plus importants que des gains statutaires ou monétaires (de Singly, 1996b). Une redéfinition de la notion d'intérêt est ainsi nécessaire pour échapper aux apories d'une lecture trop calculatrice d'un côté, ou trop enchantée de l'autre (Lordon, 2011).

Outre sa connotation en termes d'intérêts, la lecture fondée sur la notion de stratégie matrimoniale, qui suppose *a minima* un ajustement assez fort entre socialisation de groupe

3 La notion de stratégie chez P. Bourdieu vise précisément à dépasser l'opposition entre détermination sociale, perceptions intériorisées et calcul conscient : « *les stratégies matrimoniales n'ont pour principe ni la raison calculatrice ni les déterminations mécaniques de la nécessité économique, mais les dispositions inculquées par les conditions d'existence, sorte d'instinct socialement constitué qui porte à vivre comme nécessité inéluctable du devoir ou comme appel irrésistible du sentiment les exigences objectivement calculables d'une forme particulière d'économie* » (Bourdieu, 1980, p. 270). Cependant, constamment « *sur le fil du rasoir* » (Dewerpe, 1996), elle semble finalement assez difficile à manier sans retomber dans l'un ou l'autre excès.

et état du marché conjugal au moment de la mise en couple, ne semble pas la mieux à même de décrire des sociétés connaissant des transformations rapides, comme l'exode rural, la tertiarisation des emplois ou la massification scolaire. Un tel contexte est largement favorable à l'apparition d'un décalage significatif entre les *habitus* des individus et de leurs parents d'un côté, et la structure des relations entre groupes sociaux de l'autre – réduisant la cohérence d'ensemble des stratégies à peu de chose.

Si nous mobiliserons donc dans cette thèse le vocabulaire maintenant classique de « choix » du conjoint et de « marché » conjugal, c'est dans un sens bien précis qui n'implique pas l'adhésion à l'hypothèse d'une rencontre d'acteurs rationnels dans un espace de négociation abstrait de toute détermination sociale. Si nous retenons le terme de « choix », c'est dans la mesure où toutes les actions sociales résultent de décisions qui font plus ou moins l'objet d'une réflexion ou d'un calcul, fondés sur des catégories de perception et de raisonnement socialement construites. La décision de se mettre en couple avec une personne particulière, si elle n'est pas censée, dans les sociétés occidentales contemporaines, obéir à une autre logique que celle du sentiment, n'en est pas moins sujette à des évaluations similaires à celles mises en œuvre par les acteurs dans les autres domaines de la vie sociale (Lordon, 2011). Dans la mesure où il n'est pas de choix décisive – choisit-on plus sa profession que l'on choisit son conjoint ? –, ce terme nous apparaît comme le plus neutre et le plus approprié pour désigner le mécanisme par lequel les actions individuelles aboutissent aux phénomènes de l'homogamie et de l'hypergamie.

Comme toutes les actions sociales, le choix du conjoint est fortement influencé ou contraint par un grand nombre de facteurs que nous développerons ci-dessous, et dont l'un des principaux est la rencontre de conjoints potentiels correspondant au profil considéré comme acceptable par un individu – et réciproquement – dans un cadre favorable. C'est dans ce sens que M. Bozon et F. Héran (2006, p. 16) ont retenu l'expression « découverte du conjoint », se démarquant eux aussi des analyses en termes de stratégies qui prévalaient dans la littérature des années 1980. Si cette expression ne s'applique pas à notre analyse, qui n'aborde pas directement la question des lieux de rencontre, notre utilisation du concept de « marché conjugal » visera à insister comme ces auteurs sur l'importance des facteurs structurels déterminant les chances de rencontres entre individus appartenant à différents groupes sociaux. Dans notre approche, plus proche de celle des démographes (Goldman,

Westoff et Hammerslough, 1984) ou des sociologues attentifs aux effets de la structure sociale sur les interactions entre groupes (Simmel, [1908a] 1999 ; Blau, 1994), un marché correspond avant tout à la rencontre entre une offre et une demande. La rationalité des acteurs sur ce marché est d'autant plus limitée que les normes sociales actuelles dénie sa légitimité dans le champ des relations familiales, et particulièrement celui des rencontres amoureuses, que l'information y est très imparfaite, et qu'il est impossible de mettre en relation l'ensemble des candidats potentiels avant de former les couples (Oppenheimer, 1988 ; Todd et Miller, 1999). Illustrant la validité de cette approche, des simulations récentes (Billari *et al.*, 2007) fondées sur des modèles d'acteurs tenant à la fois compte de la disponibilité de conjoints socialement proches et de la pression à la mise en couple exercée par l'expérience de cet événement dans son entourage ont ainsi démontré qu'elle permet de reproduire les courbes d'âges de mise en couple effectivement observées.

Sur ce marché qui n'est pas si atypique qu'il y paraît – que l'on pense au marché du travail –, il est d'usage de distinguer trois ensembles de facteurs qui déterminent les chances de mise en couple entre individus appartenant à différents groupes sociaux (Kalmijn, 1991, 1998 ; Van Leeuwen et Maas, 2005). Premièrement, les opportunités de rencontre, qui sont déterminées par la structure du marché conjugal auquel ont accès les individus, que ce soit au niveau local ou à celui de l'ensemble d'une société. Deuxièmement, l'influence des tiers et le contrôle social sur le choix du conjoint. Troisièmement, les préférences des individus, qui sont naturellement socialement situées, et peuvent être considérées comme des normes sociales intériorisées – encore que des divergences importantes puissent exister au sein d'une société entre différents groupes sociaux ainsi qu'entre générations. Nous allons maintenant présenter les tendances de long terme des sociétés qui affectent ces différentes dimensions.

3 L'individualisation et le choix du conjoint chez les auteurs classiques

Les tendances que nous venons de présenter concernant les critères de choix du conjoint et la prégnance du contrôle social s'inscrivent dans un mouvement plus large d'individualisation et de plus grands contacts entre groupes sociaux qui a été considéré par

les auteurs classiques de la sociologie comme constitutif de la modernité. Nous commençons par présenter les théories développées par ces auteurs, avant de préciser le cadre, désigné de manière générique sous le nom de théories de la modernisation, qui est généralement retenu par les chercheurs contemporains pour l'analyse des variations de l'homogamie dans le temps et dans l'espace.

3.1 Tocqueville et l'effacement des lignées familiales

L'idée d'un déplacement des critères de choix du conjoint et de formation de la famille se rencontre déjà chez Alexis de Tocqueville. Parmi les traits caractéristiques distinguant la société d'Ancien Régime et la société démocratique symbolisée par les États-Unis, cet auteur avait déjà formulé une opposition entre choix du conjoint sur une base d'intérêts économiques et sur la base de préférences et d'une vision du monde communes :

Dans les pays aristocratiques, le mariage a plutôt pour but d'unir des biens que des personnes ; aussi arrive-t-il quelquefois que le mari y est pris à l'école et la femme en nourrice. Il n'est pas étonnant que le lien conjugal qui retient unies les fortunes des deux époux laisse leurs cœurs errer à l'aventure. Cela découle naturellement de l'esprit du contrat.

Quand, au contraire, chacun choisit toujours lui-même sa compagne, sans que rien d'extérieur le gêne, ni même le dirige, ce n'est d'ordinaire que la similitude des goûts et des idées qui rapproche l'homme et la femme ; et cette même similitude les retient et les fixe l'un à côté de l'autre.

(de Tocqueville, [1840] 1961, p. 284-285)

Cette évolution trouve pour Tocqueville son origine dans le peu d'importance qu'accordent les membres des sociétés démocratiques aux lignées familiales, qui s'effacent devant le lien conjugal. L'héritage de la terre, qui sert d'ancrage à la famille, est aussi le support de la lignée lorsque le droit d'aînesse prévaut, alors que des règles égalitaires de partage font de la famille une « chose vague, indéterminée, incertaine » (de Tocqueville, [1835] 1961, chap. III). Bien que l'auteur ne traite pas directement de la question de l'homogamie, on perçoit bien comment, dans le cadre du mouvement d'égalisation des conditions qu'il décrit, le nouvel ordre familial peut entraîner une moindre proximité des

conjoints en termes purement économiques, potentiellement au profit d'une ressemblance en termes culturels.

3.2 Durkheim et le développement de la « famille conjugale »

Les mêmes thèmes se retrouvent dans la description de la montée de la « famille conjugale » chez Émile Durkheim⁴. Le fondateur de la sociologie française voit en effet la famille reposer de moins en moins sur des enjeux économiques liés à l'héritage ou à la production de biens, ce qui renforce le rôle des liens affectifs :

Ce qui est certain, c'est que [la solidarité domestique] s'est transformée ; elle dépend de deux facteurs : les personnes et les choses. Nous tenons à notre famille parce que nous tenons aux personnes qui la composent ; mais nous y tenons aussi parce que nous ne pouvons pas nous passer des choses, et que, sous le régime du communisme familial, c'est elle qui les possède. De l'ébranlement du communisme, il résulte que les choses cessent de plus en plus d'être un ciment de la société domestique. La solidarité domestique devient toute personnelle. Nous ne sommes attachés à notre famille que parce que nous sommes attachés à la personne de notre père, de notre mère, de notre femme, de nos enfants. Il en était tout autrement autrefois où les liens qui dérivait des choses primaient au contraire ceux qui venaient des personnes, où toute l'organisation familiale avait avant tout pour objet de maintenir dans la famille les biens domestiques, et où toutes les considérations personnelles paraissaient secondaires à côté de celles-là.

(Durkheim, [1892] 1975, p. 44)

Comme chez Tocqueville, on peut penser que la moindre importance accordée aux facteurs matériels autorise une plus grande liberté de choix du conjoint pour les jeunes. L'un des apports de Durkheim consiste à souligner le rôle que joue le développement de l'État joue dans l'évolution de la famille. Il participe en effet à affaiblir l'autorité du père en la limitant dans le temps (avec l'émancipation des enfants à leur majorité) comme dans son champ d'action (contrôles de l'Assistance publique). En outre, en développant la scolarisation et en taxant les héritages, il donne plus d'importance au mérite scolaire dans la

4 On trouve aussi chez Durkheim l'idée d'une évolution linéaire des structures familiales, dans le sens d'une contraction progressive menant à la famille conjugale. Cette théorie a été clairement invalidée par les recherches réalisées depuis (voir Todd, 2011 pour une somme récente). De même, nous n'aborderons pas ici les considérations normatives discutables formulées à propos des modèles familiaux par les différents auteurs, qui ne concernent pas directement les déterminants du choix du conjoint.

détermination du statut social, émancipant de ce fait les individus de leurs origines familiales.

Durkheim développe plus largement une théorie de la différenciation croissante des sociétés qui concerne de manière assez directe la question de la liberté du choix du conjoint. Ainsi qu'il l'explique dans *La Division du travail social* ([1893] 2013), l'augmentation de la « *densité matérielle et morale* » des sociétés, qui découle de l'augmentation de la densité de population, de l'urbanisation et du développement des transports et des communications, entraîne une division de plus en plus forte du travail et une spécialisation croissante des activités. Les grandes villes offrent le meilleur exemple de cette évolution, mais le développement des transports et des moyens de communication étend ses conséquences à l'ensemble du territoire.

Suite à cette évolution, les appartenances de groupe peuvent de moins en moins se fonder sur la similarité entre individus (solidarité mécanique), mais doivent découler de leur interdépendance (solidarité organique). Les groupes étant moins homogènes, la mobilité sociale s'accroît, la conscience commune qui lie les individus entre eux s'affaiblit et la tradition perd en influence :

[...] les mêmes causes qui, d'une manière générale, allègent le joug collectif, produisent leur effet libérateur à l'intérieur de la corporation comme au-dehors. À mesure que les organes segmentaires fusionnent, chaque organe social devient plus volumineux, et cela d'autant plus que, en principe, le volume total de la société s'accroît au même moment. Les pratiques communes au groupe professionnel deviennent donc plus générales et plus abstraites, comme celles qui sont communes à toute la société et, par suite, elles laissent la place plus libre aux divergences particulières. De même, l'indépendance plus grande dont les générations nouvelles jouissent par rapport à leurs aînées ne peut manquer d'affaiblir le traditionalisme de la profession ; ce qui rend l'individu encore plus libre d'innover.

(Durkheim, [1893] 2013, sect. II.III.IV)

En prolongeant l'interprétation durkheimienne, on peut avancer qu'emprise familiale moindre et affaiblissement de la tradition laissent ainsi l'individu libre d'« *innover* » entre autres matière de choix du conjoint. C'est cette lecture que retiennent les théories de la modernisation.

3.3 Simmel et l'entrecroisement des cercles sociaux

Le progrès de l'individualisme est théorisé par Georg Simmel sous une forme légèrement différente, mais qui rejoint en grande partie la vision durkheimienne. Pour cet auteur, l'« *accroissement quantitatif du groupe* », dû par exemple à l'unification nationale et à l'ouverture internationale des marchés pour ce qui concerne les corporations professionnelles, est associée à une plus grande différenciation sociale, une plus grande hétérogénéité interne des groupes sociaux, et des frontières entre groupes moins nettes (Simmel, [1908b] 1999). En effet, un groupe large laisse plus de champ à l'expression de l'individualité de ses membres. Du fait de l'importance du critère quantitatif, la grande ville apparaît, comme chez Durkheim, comme le lieu où s'exprime le plus nettement la modernité.

Plutôt que d'opposer solidarités mécanique et organique, Simmel recourt au concept de cercles sociaux, qui désignent les multiples groupes d'appartenance des individus. Dans l'état de communauté, les cercles se superposent largement : les individus présentent tous les mêmes affiliations, de la plus locale (le village) à la plus éloignée (la nation). Aussi, les individus sont fortement enserrés dans le groupe, et leur conscience est largement dépendante du collectif. Au contraire, dans l'état de société, les individus appartiennent à un plus grand nombre de cercles qui se recoupent assez peu. La personnalité s'affirme puisqu'« *elle cesse d'être solidement enracinée dans un seul cercle pour se situer au croisement de nombreux cercles* » (Simmel, [1908a] 1999, p. 417). L'individu, divisé en multiples appartenances indépendantes, émerge comme une instance supérieure à la somme de ses rôles sociaux.

Cette lecture a amené Simmel à poser les jalons d'une analyse des déterminants structurels de l'homogamie qui a été reprise plus tard par Peter Blau et Joseph Schwartz (1984). Le sociologue allemand remarquait en effet à propos des cercles sociaux que « *plus ils sont nombreux, moins il y a de chances que d'autres personnes présentent la même combinaison de groupes, que ces nombreux cercles se recoupent encore en un autre point* » (Simmel, [1908a] 1999, p. 416). En d'autres termes, un accroissement de la différenciation sociale rend chaque individu de plus en plus particulier, diminuant la probabilité de

rencontrer son parfait équivalent structural (selon le concept développé dans le champ de l'analyse des réseaux sociaux).

Dans le domaine du choix du conjoint, ce mécanisme permet de conclure qu'un fort entrecroisement des cercles sociaux rend de moins en moins probable la formation d'un couple parfaitement homogame sur tous les tableaux. La recherche d'un partenaire socialement proche, autrefois aisée, devient problématique, en dépit de l'élargissement de l'« *aire de recrutement du conjoint* » (selon l'expression de Bozon et Héran, 1987a) :

Aux époques de séparations strictes en clans, en familles, en états définis par la profession ou la naissance, en castes, etc., par comparaison avec les sociétés avancées ou libérales on ne dispose que d'un cercle relativement réduit où trouver son conjoint, homme ou femme. Mais pour autant que nous puissions appréhender ces situations et en juger par certaines analogies avec notre temps, le choix n'était pas difficile, du fait des individus : la différenciation réduite des personnes et des relations conjugales impliquait que chaque homme, en cas de compatibilité extérieure, pouvait s'arranger avec presque toute jeune fille du cercle possible sans, des deux côtés, directives ou exclusives psychologiques vraiment spécifiques à l'individu. Le progrès de la civilisation a bouleversé cette situation des deux côtés. Le cercle où l'on peut choisir un conjoint s'est énormément accru du fait du mélange des conditions, de l'élimination des barrières religieuses, de l'affaiblissement de l'autorité parentale, de la liberté de mouvement tant au sens géographique qu'au sens social, etc. Mais en revanche la sélection individuelle est bien plus stricte et relève en droit et en fait de l'inclination la plus personnelle, la conscience que parmi tous les êtres ces deux-là sont « faits » l'un pour l'autre et seulement l'un pour l'autre – évolution encore inouïe même pour la bourgeoisie du XVIII^e siècle.

(Simmel, [1908b] 1999, p. 700)

On passe donc d'une situation dans laquelle l'homogamie était automatique à une situation dans laquelle elle est problématique : le choix doit faire l'objet d'une recherche et d'une découverte, les individus sont censés « trouver » le partenaire qui leur correspond. Or, dans un contexte de fort entrecroisement des appartenances, des hiérarchies contradictoires tendent à se développer : un individu au statut social élevé sur une dimension peut se trouver plus bas sur une autre (phénomène aussi désigné sous le nom d'incohérence de statut d'après Lenski, 1954). Même si les individus cherchent un conjoint qui soit leur parfait double social, ce qui revient à former un couple fortement homogame, ils se voient forcés de

choisir de s'assortir selon une dimension plutôt qu'une autre. On voit l'importance que peut prendre cette observation dans une période de massification scolaire et d'accroissement de la mobilité sociale : origine sociale, diplôme et profession ont des chances de ne pas toujours coïncider ; les individus auront donc à choisir entre caractéristiques héritées et caractéristiques acquises, ces dernières se divisant à leur tour entre aspect culturel et aspect économique (van Leeuwen et Maas, 2005, p. 18). Ainsi, en étendant l'analyse de Simmel, on peut s'attendre à ce que l'homogamie s'affaiblisse au moins selon certaines dimensions lorsque l'entrecroisement des appartenances augmente, que ce soit au cours du temps, ou d'une société à une autre.

4 Les évolutions récentes des sociétés européennes

Ces tendances de long terme identifiées dès les origines de la sociologie prennent, dans les sociétés contemporaines qui constituent l'objet de notre étude, la forme de processus plus précis qui concernent à la fois le domaine de la famille et celui de la stratification sociale. Certaines ne sont que le prolongement d'évolutions qui couvrent l'ensemble du XX^e siècle, alors que d'autres sont apparues dans la seconde moitié du siècle et réorientent de manière notable les tendances précédentes.

4.1 Apogée et déclin du modèle conjugal de l'après-guerre

Les transformations de la famille qui ont pris place depuis cette période sont bien connues (Segalen, [1981] 2010, part. II ; de Singly, [1993] 2014 ; Widmer, 2014). La famille conjugale décrite par Durkheim et théorisée plus tard par la sociologie structuro-fonctionnaliste américaine connaît son apogée dans l'après-guerre, avec un mariage précoce, un célibat peu fréquent, des divorces rares et une natalité élevée. La famille nucléaire composée du couple parental et de ses enfants avec une forte division sexuelle des rôles constitue alors le modèle dominant – même si la réalité est toujours demeurée plus diverse que les théories sociologiques de l'époque ne la voyaient. Dans la continuité des auteurs classiques, Talcott Parsons (Parsons et Bales, 1955) a particulièrement insisté sur l'isolement de la famille nucléaire par rapport à sa parenté, lié à la néolocalité et à la forte mobilité géographique de la société américaine. Cette affirmation ne doit cependant pas

conduire à conclure que – pour l’auteur comme dans la réalité – les relations familiales extérieures au couple disparaissent complètement. L’évolution, pour réelle quelle soit, est moins radicale : d’une part, les ménages ont toujours été de taille réduite, même s’ils s’imbriquaient dans une parenté plus large (Goody, 2001, p. 211) ; d’autre part, les relations entre parents et enfants après la décohabitation demeurent intenses encore aujourd’hui (Bonvalet et Lelièvre, 2012, p. 59). Ce qui caractérise la famille nucléaire moderne, c’est avant tout que les conjoints se choisissent indépendamment des influences de leurs familles élargies, ne dépendent pas d’elles économiquement et ne définissent pas principalement leurs identités par rapport à elles. Dans un tel cadre, le maintien des relations avec la famille étendue peut même contribuer à l’autonomie de l’individu (Déchaux, 1999, p. 30).

Cette conjoncture exceptionnelle à l’échelle historique s’effrite dans les pays occidentaux dès le milieu des années 1960, en partant de l’Europe du Nord. On assiste alors à un développement de la diversité des formes familiales, mouvement que l’on peut qualifier à partir de trois tendances (Widmer, 2014) : privatisation (ou désinstitutionnalisation), individualisation et affaiblissement de la séparation des rôles de genre. Cette évolution a pu être lue comme l’entrée dans une troisième phase historique qui, après une lutte séculaire entre amour et mariage, puis une phase dans laquelle l’un et l’autre se renforçaient mutuellement, a vu le premier remettre en cause le second au nom de la primauté des sentiments en tant que fondation du couple (de Singly, [1993] 2014, sect. 3.1).

4.2 L’affaiblissement du mariage et la montée des cohabitations

Parmi les nombreuses évolutions des modèles familiaux, le développement des cohabitations hors mariage et l’augmentation des séparations et des divorces (Pailhé *et al.*, 2014) sont les plus directement liées aux modes de mise en couple et donc au choix du conjoint. Alors que pour les générations nées avant la Seconde Guerre mondiale, la mise en ménage intervenait en même temps que le mariage et constituait un engagement durable voire définitif, la mise en couple devient pour les générations suivantes un processus plus progressif et réversible. Cette évolution s’inscrit dans une tendance plus large à l’allongement de la jeunesse et au déclin des rites de passage (Galland, [1991] 2011, chap. 3) : l’indétermination sociale caractéristique de la jeunesse s’accommode bien de la forme souple de vie en couple que représente la cohabitation. Elle peut être lue comme le signe

d'une désinstitutionnalisation de la vie en couple, sanctionnant une moindre prégnance des normes sociales et une moindre implication de l'entourage dans le choix du conjoint. Ainsi le mariage perd-il peu à peu son caractère de rite de passage : intervenant de plus en plus fréquemment alors que le couple a déjà un ou plusieurs enfants, il ne constitue qu'une étape dans la vie du couple, implique moins la parenté et donne lieu à des fêtes moins éclatantes et avec moins d'invités (Bozon, 1988).

Ce mouvement d'ensemble pourrait impliquer une moindre détermination sociale du choix du conjoint, et donc un affaiblissement de l'homogamie. Cependant, les effets exacts du développement des cohabitations hors mariage sur le choix du conjoint sont discutés (Hamplová, 2009). Pour certains auteurs, la possibilité de réaliser un mariage « à l'essai » (Roussel, 1978) renforce en la prolongeant la sélection à l'œuvre dans le processus de formation des couples depuis la rencontre jusqu'au mariage : les couples mal assortis ayant plus tendance à se séparer (un résultat attesté, quoique d'ampleur modeste, sur lequel nous reviendrons au chapitre 1), seules les unions les plus homogames finiraient par donner lieu à un mariage.

Mais cette lecture, qui compare couples mariés et couples cohabitants à un instant donné, ne répond pas véritablement à la question de l'effet du développement des cohabitations au cours du temps sur l'intensité de l'homogamie dans une société. De ce point de vue, nous chercherons à montrer qu'un autre scénario est à l'œuvre, qui n'est pas incompatible avec le constat d'une plus forte homogamie des couples mariés. On peut en effet penser que le caractère progressif de la mise en couple diminue les risques associés à l'aventure que constitue la formation d'une union hétérogame. Lorsque l'engagement est définitif et qu'il n'est pas précédé d'une période de vie commune, le risque de tensions liées à des styles de vie ou à des attentes trop différentes de la part des deux conjoints est particulièrement élevé⁵. Au contraire, lorsque la frontière entre le flirt et le mariage s'estompe, laissant place à une série d'étapes progressives menant de l'engagement le plus faible à l'engagement plus fort, les unions socialement improbables ont une chance de perdurer. Soulignons que cette hypothèse n'est pas incompatible avec la précédente : à un instant donné, les couples mariés peuvent présenter une plus grande proximité sociale ;

5 Pour un exemple de processus de négociation concernant les normes de répartition des tâches ménagères, cf. Jean-Claude Kaufmann (1992).

mais du point de vue des variations temporelles ou spatiales, le développement des cohabitations peut néanmoins avoir pour effet de diminuer l'homogamie à la fois des couples cohabitants et des couples mariés.

La proximité sociale des deux conjoints n'est pas seule en cause : le sens de l'écart de position sociale entre homme et femme (hypergamie ou hypogamie) peut aussi être affecté par le développement des cohabitations. L'institution du mariage implique en effet une inscription dans un modèle relativement bien défini du couple, avec un engagement de long terme, ainsi qu'un certain nombre de garanties pour les époux, et en particulier pour l'épouse. Ces assurances, comme celle du versement d'une pension en cas de divorce ou de décès du mari, peuvent favoriser la spécialisation des rôles au sein du couple, c'est-à-dire l'investissement dans la carrière professionnelle de l'homme, celle de la femme étant secondaire ou inexistante (de Singly, 1987b ; Nicole-Drancourt, 1989).

Même si son sens peut varier selon les pays, l'institution du mariage est donc susceptible d'apparaître aux femmes qui cherchent à échapper au moins partiellement aux rôles de genre comme le symbole de la division sexuée du travail au sein du couple. Ainsi, Françoise Battagliola (1988) observait dans le cas français que la remise en cause du mariage était d'abord le fait des femmes diplômées qui craignaient de perdre leur indépendance en se mariant avant d'avoir terminé leur insertion professionnelle. Michel Bozon (1990, p. 592) notait aussi que les femmes qui vivaient en cohabitation hors mariage insistaient moins sur la supériorité professionnelle de l'homme comme critère d'appréciation du conjoint. Au total, si le développement des cohabitations doit entraîner une diminution de l'homogamie, on peut donc aussi s'attendre à ce qu'il soit associé à une diminution de la propension à l'hypergamie féminine.

4.3 L'élévation du statut des femmes

Plus encore que le développement des cohabitations, l'élévation du statut des femmes depuis les années 1960 est à la fois le signe et la cause d'une évolution des normes concernant le choix du conjoint. Plusieurs travaux ont ainsi montré une convergence des préférences des hommes et des femmes en matière conjugale (Bozon, 1990 ; Buss *et al.*,

2001) : les premiers insistent plus qu'avant sur les aspects tenant au statut social, les secondes plus sur les dimensions esthétiques et relationnelles.

Si cette évolution semble devoir se traduire par une diminution de l'hypergamie féminine, son effet sur l'homogamie a fait l'objet de débats. En effet, des travaux (Mare, 1991 ; Blossfeld et Timm, 2003a) ont souligné que la plus grande prise en compte par les hommes du statut social de leurs conjointes potentielles pouvait entraîner un renforcement de l'homogamie. On assisterait alors en quelque sorte à un transfert de l'hypergamie vers l'homogamie ; la diminution des asymétries genrées aurait pour corollaire une diminution des écarts de statut entre conjoints : égalité mais proximité. Cette lecture nous paraît cependant incomplète : elle ne tient compte que de l'aspect économique du choix du conjoint, et elle ignore l'évolution des préférences des femmes.

Tout d'abord, même dans un modèle où l'activité professionnelle des femmes est réduite, leur statut social constitue un critère majeur de choix, pour des raisons à la fois de proximité culturelle et de reproduction sociale – *via* par exemple la représentation sociale du ménage ou l'éducation des enfants. Ensuite, la forte dépendance des femmes par rapport aux revenus de leur conjoint fait de la mise en couple un enjeu économique et statutaire majeur pour elles. Au contraire, lorsque l'activité des femmes est développée, et le statut social et les revenus afférents relativement élevés, les femmes peuvent accorder plus d'importance à d'autres dimensions. C'est bien ce que les travaux portant sur l'évolution des préférences féminines ont relevé. De fait, nous verrons que l'augmentation de l'activité féminine ne se traduit pas dans nos résultats par un renforcement de l'homogamie.

4.4 L'évolution des modèles conjugaux et la « relation pure »

L'évolution des modèles conjugaux, évidemment liée à celle du statut des femmes, peut renforcer les tendances que nous venons de commenter. Le passage vers un modèle de moins en moins instrumental et de plus en plus affectif a pu être vu par certains auteurs comme porteur d'une plus grande égalité entre homme et femme au sein du couple. L'amour serait ainsi à l'origine d'une suspension de la domination masculine (Bourdieu, 1998, p. 146). La vision quelque peu désincarnée de la « *relation pure* » (Giddens, 2004) a été justement critiquée pour son manque d'attention aux inégalités de genre persistantes

(Jamieson, 1999). Sans verser dans un optimisme sans limites, il nous paraît cependant que l'accent mis sur l'échange entre conjoints et la construction d'un univers commun, qui n'est pas un phénomène absolument nouveau, peut participer à la réduction des asymétries entre hommes et femmes dans le choix du conjoint. En effet, l'hypergamie féminine n'a de sens que dans le cadre d'un modèle inégalitaire du couple, supposant des critères d'appréciation très différents chez les deux sexes. Lorsque les deux conjoints sont censés contribuer chacun à la construction de l'identité personnelle de l'autre, cette asymétrie perd de sa signification.

L'insistance croissante sur le partage d'une vision commune au sein du couple peut elle aussi faire anticiper un renforcement de l'homogamie, du moins du point de vue culturel, c'est-à-dire du niveau d'éducation des conjoints⁶. Mais ici encore, d'autres auteurs (Alberoni, [1979] 1993, p. 25, 42) ont à l'inverse avancé que l'amour ne pouvait naître qu'en présence d'obstacles, et en particulier de barrières liées à des différences sociales : la passion se nourrit de la différence. Pour cela, Ulrich Beck et Elizabeth Beck-Gernsheim (1995, p. 195) affirment que la montée en puissance du sentiment amoureux dans la mise en couple entraîne non seulement plus d'égalité entre hommes et femmes au sein du couple, mais aussi une moins grande proximité sociale des conjoints. Cette vision doit évidemment être mise à l'épreuve, ne serait-ce que parce qu'il est loin d'être vérifié que l'ensemble des couples, ou même leur majorité, se forment sur un mode passionnel allant radicalement à l'encontre des appartenances sociales. Les rencontres sous forme de « coup de foudre », qui correspondent potentiellement le plus à ce profil, sont ainsi nettement minoritaires (Bozon et Héran, 1987b, p. 966).

4.5 L'évolution de la stratification sociale

Les structures des sociétés européennes ont elles aussi considérablement évolué dans la période récente, et diffèrent en outre considérablement les unes des autres dans les années récentes (Breen, 2004 ; Róbert, 2010). Trois dimensions majeures peuvent être distinguées, qui peuvent agir sur différents aspects du choix du conjoint. Nous les résumons ici rapidement, avant de les observer plus en détail dans les chapitres qui suivent.

6 Au contraire, la profession est souvent considérée comme relevant exclusivement du domaine économique. Cette assimilation nous paraît abusive, tant les univers professionnels renvoient à des univers culturels parfois très contrastés (un exemple classique en la matière étant fourni par Bourdieu, 1979).

La première tient à la variation de la proportion de la population agricole, qui a très fortement diminué, quoique à des rythmes très différents, dans les pays européens. Cette évolution est encore visible en France au cours des quarante dernières années. En effet, ce pays a connu un déclin relativement tardif de sa population agricole, qui représentait encore un quart des ménages en 1945, et 15 % en 1968 (Molinier, 1977). Encore aujourd'hui, des différences importantes entre pays persistent en Europe, la Roumanie, la Pologne, la Grèce ou encore le Portugal comprenant encore des populations agricoles particulièrement importantes (allant de 10 % à 40 %). Or, les agriculteurs figurent parmi les groupes à l'homogamie la plus marquée ; la disparition du monde paysan participe donc à abaisser l'homogamie à la fois du fait de la diminution des effectifs de ce groupe, mais aussi en affaiblissant son identité et sa cohésion (Bourdieu, 1962). Plus largement, la diminution de la proportion de petits indépendants prolonge le mouvement de long terme liant développement du salariat et perte d'importance des lignées du fait du déclin de l'héritage des moyens de production. Pour les travailleurs indépendants et leurs enfants qui envisagent de reprendre l'entreprise familiale, le choix d'un conjoint est aussi souvent celui d'un collaborateur – ou du moins, dans les années récentes, d'un soutien essentiel –, ce qui peut entraîner des tensions avec les beaux-parents (Bertaux-Wiame, 2004 ; Bessière, 2014).

Une seconde tendance majeure tient à la tertiarisation de l'économie et au recul de l'emploi ouvrier au profit d'emplois en col blanc, parfois à statut élevé (cadres, professions libérales et intellectuelles, ou professions intermédiaires), mais aussi le plus souvent au niveau d'employé, qualifié ou non. L'expansion des catégories supérieures et la mobilité sociale ascendante qui en découle peuvent entraîner une augmentation de l'entrecroisement des appartenances et un affaiblissement de la cohérence de ce groupe, et ainsi diminuer sa tendance à l'entre-soi. De son côté, l'évolution d'emplois manuels vers des emplois non manuels parmi les classes populaires, associé au déclin de la grande industrie au profit d'établissements de taille plus restreinte, est à l'origine d'un effritement de la classe ouvrière (Mauger, 2006). Là encore, des écarts importants sont visibles de ce point de vue entre pays européens, les employés comme les professions intermédiaires et supérieures représentant une proportion de la population bien plus élevée dans l'Europe du Nord et du Nord-Ouest que dans l'Europe du Sud et de l'Est.

La massification scolaire est la troisième tendance à l'œuvre. Comme le développement des emplois au sein des classes supérieures, la croissance de la part de diplômés du supérieur augmente l'hétérogénéité de ce groupe, et diminue potentiellement son homogamie. L'augmentation générale du niveau d'éducation entraîne une certaine unification culturelle de la population, et en particulier de la jeunesse (Van Leeuwen et Maas, 2005, p. 6 ; Galland, [1991] 2011, part. 3). En revanche, dans la logique de la citation de P. Bourdieu (Bourdieu, 1974, p. 74) présentée plus haut, des travaux (Mare, 1991 ; Blossfeld et Timm, 2003b) ont montré que l'allongement des scolarités renforçait l'homogamie en augmentant la fréquence des rencontres sur le lieu d'études. Nous montrerons cependant que cet effet, s'il est bien réel, n'est pas assez fort pour annuler celui de l'unification culturelle.

Comme pour les autres dimensions, des différences importantes en termes de niveau d'éducation existent entre pays européens (Müller et Kogan, 2010). Les pays d'Europe du Sud et les pays de l'ancien bloc soviétique présentent des taux de diplômés du supérieur généralement faibles par rapport au Nord-Ouest du continent ; en revanche, la proportion de diplômés du secondaire complet (lycée) est relativement élevée par rapport au reste du continent dans les pays ayant connu un système d'éducation socialiste. Outre ces différences de niveau, le degré de *différenciation* entre filières, ainsi que la position des paliers d'orientation varient assez fortement, et introduisent des différences dans la sélectivité sociale des diplômes. L'opposition la plus nette de ce point de vue met en regard l'Allemagne et certains de ses voisins, qui orientent précocement les élèves entre filières générale et professionnelle, et les pays scandinaves ou d'Europe du Nord (mais aussi parfois du Sud), qui ne proposent qu'une filière commune jusqu'à l'entrée dans le second cycle du secondaire (lycée). Nous verrons que ces variations ont des effets tant sur le niveau que sur la structure de l'homogamie.

5 Le cadre d'analyse contemporain du choix du conjoint

5.1 La notion d'ouverture sociale

Dans les années récentes, l'homogamie a le plus souvent été étudiée sous l'angle de la question de l'*ouverture sociale*. Ce concept, intimement lié à celui de mobilité sociale dans ses divers aspects, désigne la mesure dans laquelle les groupes sociaux sont plus ou moins étanches les uns aux autres (Ultee, 2007). Cette approche d'inspiration wébérienne considère les mouvements d'individus d'un groupe social à l'autre, ou les contacts entre groupes liés à la formation des couples (*connubium*) ou aux relations amicales (*convivium*), comme des indices de la structure sociale.

Parmi les premiers chercheurs à avoir porté leur attention sur l'homogamie du point de vue de la stratification sociale, Seymour Martin Lipset et Hans Zetterberg (1959) ont considéré l'ouverture sociale comme un facteur de cohésion d'une société. À la différence de la solidarité organique théorisée par Durkheim, qui repose sur une complémentarité des individus ou des groupes sociaux du point de vue de la division du travail, cette cohésion sociale tient à l'existence de relations concrètes entre les différents groupes. Ainsi, la mobilité sociale et l'hétérogamie amènent des relations familiales à franchir les frontières entre groupes sociaux et peuvent renforcer le sentiment d'une appartenance commune. Cependant, les auteurs, formulant ce qui deviendrait l'hypothèse dite Lipset-Zetterberg, concluaient sur la base des données dont ils disposaient que les pays occidentaux différaient peu en termes de degré d'ouverture sociale. Cette conclusion provisoire a finalement été rejetée par des travaux ultérieurs du point de vue de l'homogamie comme de celui de la mobilité sociale.

L'existence de variations entre pays acquises, une première question des études comparatives de la stratification sociale portait donc sur l'existence ou non d'un lien entre mobilité sociale intergénérationnelle et homogamie. En effet, ce n'est qu'à cette condition qu'il est possible de considérer ces deux caractéristiques comme des indices du phénomène plus large de l'ouverture sociale d'une société. Nous avons déjà développé des éléments théoriques permettant d'anticiper l'existence d'une telle relation : une forte mobilité sociale est à l'origine d'un entrecroisement des appartenances qui doit au minimum diminuer

l'homogamie en termes d'origine sociale, mais aussi potentiellement celle liée à d'autres critères. Comme nous le verrons, cette relation est nettement attestée, validant le recours à la notion d'ouverture sociale. Cependant, les études ont conclu que l'association entre niveaux d'éducation des conjoints (homogamie) était supérieure à celle entre les individus et leurs parents (hérédité sociale), faisant de l'homogamie un meilleur indicateur de la structure sociale que la mobilité intergénérationnelle, pourtant traditionnellement privilégiée dans ce champ de recherche (Ultee et Luijkx, 1990 ; Katrňák, Fučík et Luijkx, 2012).

Une seconde question centrale de la sociologie de la stratification a porté sur la relation entre l'ouverture d'une société et le niveau des inégalités qui y règnent. Ce lien s'avère plus difficile à établir, et fait encore l'objet de débats et de recherches. La réponse peut en particulier dépendre du champ géographique sur lequel se fondent les comparaisons internationales : on peut penser qu'une corrélation a d'autant plus de chances d'être relevée que le champ retenu est large, et que les différences dans le niveau des inégalités sont par conséquent amples. Dans le cas du seul continent européen, les écarts sont relativement réduits, et les études existantes – ainsi que nos propres analyses ci-dessous – aboutissent à des résultats ambigus (Erikson et Goldthorpe, 1992 ; Breen, 2004).

Ce constat ne nous semble pas complètement inattendu. Sur le plan théorique, en effet, l'interprétation de l'ouverture sociale comme un indice de la cohésion d'une société peut mener à des conclusions contradictoires concernant le lien entre ouverture sociale et inégalités. Selon une première lecture, une société dans laquelle les contacts entre groupes sociaux sont fréquents et intenses pourrait faire preuve d'une moindre tolérance aux inégalités entre groupes. En prenant la causalité dans l'autre sens, dans une société égalitaire, les barrières entre groupes sont moins marquées, et les échanges plus fréquents. Au contraire, dans une société inégalitaire, les individus appartenant aux groupes les plus favorisés ont beaucoup à perdre à s'allier avec des individus de statut inférieur, donnant tout son sens à la notion de mésalliance. En outre, les écarts de revenus entraînent des différences de styles de vie, de lieux fréquentés et de goûts : les ressources économiques sont en effet un facteur discriminant peu étudié, mais central, pour toute une série de pratiques culturelles coûteuses – consommation, loisirs, sorties ou encore vacances (Coulangeon, Menger et Roharik, 2002 ; Yaish et Katz-Gerro, 2010). Si l'on admet que l'apparition du

sentiment amoureux, et plus largement d'affinités entre individus, dépend non seulement de l'existence de contacts, mais aussi du partage de goûts, de manières et d'intérêts communs (Bourdieu, 1979, p. 267 ; DiMaggio et Mohr, 1985 ; Bozon, 1991), on doit s'attendre à observer un lien positif entre inégalités économiques et homogamie.

Mais une seconde lecture, attribuant plus d'importance au conflit social, dans une logique aussi bien wébérienne que marxiste, pourrait défendre la position inverse. La conscience de classe ne pourrait alors se développer pleinement qu'au sein des sociétés dans lesquelles les groupes sociaux sont les plus clairement séparés ; c'est elle qui serait à l'origine, *via* le renforcement de la conscience de classe et le conflit social, d'une réduction des inégalités. Cette lecture sous-tend par exemple les analyses de Werner Sombart ([1906] 1992), qui attribuent – parmi de nombreux autres facteurs – l'absence de développement du socialisme aux États-Unis au caractère moins marqué des frontières entre classes, en particulier en termes de mobilité sociale, par rapport à l'Europe au début du XX^e siècle. Elle se retrouve aujourd'hui dans la « *spirale des classes sociales* » proposée par Louis Chauvel (2001) pour expliquer la relation historiquement négative entre inégalités et conscience de classe. La cohésion sociale ne serait alors qu'une mystification qui permettrait aux classes les plus favorisées d'accumuler les richesses sans que les autres strates de la société ne réagissent. Ce débat est, on s'en doute, difficile à trancher rigoureusement, d'autant qu'on ne peut exclure qu'un mécanisme soit valable pour décrire des différences entre pays à un moment donné, et qu'un autre s'applique aux évolutions temporelles au sein d'une société donnée : le premier renverrait à des différences politiques et culturelles, quand le second recouvrirait une dynamique plus conjoncturelle.

5.2 Théories de la modernisation et évolutions temporelles

Un second axe de recherche, difficilement dissociable de la question de l'ouverture sociale, vise à mesurer l'évolution temporelle de l'homogamie. Le cadre généralement retenu est celui des théories de la modernisation (Goode, [1964] 1982, chap. 10), et plus particulièrement des théories de l'industrialisation (Kerr *et al.*, 1960 ; Treiman, 1970), qui dérivent en grande partie des interprétations élaborées par les auteurs classiques déjà présentées plus haut. Ce cadre très large, dont nous avons déjà présenté les principales lignes, peut être résumé.

Les recherches récentes ont plus précisément porté sur deux questions centrales : celle des évolutions temporelles de l'homogamie, et celle, très directement liée, de son évolution avec le développement économique. Un schéma a particulièrement fait l'objet de débats. Proposé par Jeroen Smits, Wout Ultee et Jan Lammers (1998) à partir d'une étude empirique portant sur l'homogamie d'éducation, et mis à l'épreuve par plusieurs études, il s'articule en trois étapes. En se fondant sur les théories de l'industrialisation, les auteurs avancent premièrement qu'avec le développement des organisations bureaucratiques et l'augmentation de la mobilité sociale, l'importance croissante prise par l'éducation par rapport aux identités héritées telles que les origines sociales, ethniques ou géographiques implique que les individus chercheront un partenaire aussi diplômé que possible ; l'allongement des études augmente aussi la probabilité de rencontrer son conjoint dans ce cadre : tous ces facteurs devraient conduire à une augmentation de l'homogamie d'éducation. Mais les théories de la modernisation avancent aussi qu'avec la modernité, le choix du conjoint s'émancipe du contrôle des parents et l'amour prend de plus en plus d'importance ; les contacts entre groupes se développent et les inégalités diminuent : l'homogamie d'éducation devrait donc s'affaiblir. Une troisième proposition vient donc réconcilier les deux hypothèses en postulant qu'elles s'appliquent à deux phases différentes de la modernisation : l'homogamie d'éducation va se renforcer avec l'industrialisation et l'augmentation de l'importance des qualifications sur le marché du travail ; puis décroître avec le développement de « l'amour romantique » au cours de la phase de modernité tardive que connaissent aujourd'hui les pays développés, suivant les arguments classiques des théories de la modernisation. L'amour apparaît ainsi comme un luxe permis par un niveau de vie élevé et une protection assurée par l'État social. Une version révisée de ce scénario (Smits, Ultee et Lammers, 2000) avance en outre qu'une phase de saturation sera finalement atteinte, dans laquelle l'hétérogamie aurait atteint un plafond et se stabiliserait.

Ce schéma théorique assez élaboré a pour faiblesse d'avoir été élaboré à partir d'une base empirique relativement étroite, tenant à la relation en U inversé entre niveau de développement et homogamie identifiée par les auteurs. Cette relation, dont on peut douter du caractère systématique, a été critiquée par les travaux ultérieurs sur lesquels nous reviendrons en détail dans notre dernière partie. Cette élaboration *ad hoc* l'expose aussi à des critiques plus fondamentales sur le plan théorique. Il n'est en effet pas évident que

l'augmentation de l'importance des qualifications scolaires ait précédé l'essor de l'amour en tant que déterminant principal du choix du conjoint. Ces deux phénomènes sont intimement liés dans le processus de modernisation, l'importance accordée à l'amour pouvant être lue comme la conséquence de l'individualisation et de la prépondérance du capital culturel sur le capital économique (de Singly, 1996a). En outre, les travaux présentés plus haut, qui font démarrer les évolutions des modèles conjugaux bien avant le XIX^e siècle, ne nous donnent aucune raison de penser que l'amour ne gagne en importance pour le choix du conjoint qu'après que l'éducation soit devenue un déterminant majeur de la stratification sociale.

Ce débat important restera en simple toile de fond de cette thèse : du fait de notre champ d'analyse, limité à des pays développés, nous ne pourrions pas chercher à évaluer la validité de ce schéma en ce qui concerne la première phase du développement. Nos résultats confirment en tout cas l'hypothèse d'une décroissance de l'homogamie avec le développement parmi les pays étudiés, et ne soutiennent pas l'idée d'une saturation au-delà d'un certain niveau. À défaut de pouvoir contribuer au test de cette théorie d'une évolution en deux temps, nous approfondirons le rôle des multiples déterminants du choix du conjoint identifiés dans la littérature, et qui sont susceptibles de jouer un rôle plus ou moins important à différentes étapes du développement d'un pays.

5.3 Les quatre générations d'études de l'homogamie

Plusieurs auteurs (Uunk, 1996 ; Smeenk, 1998) ont distingué quatre générations successives d'études de l'homogamie, qui correspondent de près à celles des études de la stratification et de la mobilité sociale (Ganzeboom, Treiman et Ultee, 1991 ; Treiman et Ganzeboom, 2000). Ces générations correspondent autant à des périodes temporelles qu'à des combinaisons de questions de recherche et de méthodes d'analyse. Avant de les décrire, soulignons qu'elles ne couvrent que les études portant sur les implications en termes de stratification sociale du choix du conjoint. En particulier, les travaux classiques français d'A. Girard (1964) ainsi que de M. Bozon et F. Héran (2006), qui se concentrent sur l'étude des lieux et les modalités des rencontres, échappent en grande partie à ce schéma.

La première génération est apparue au tournant du XX^e siècle. Les travaux de cette génération recouraient à des statistiques descriptives simples, comme des tables

d'homogamie croisant les professions ou les niveaux d'éducation des conjoints, exprimés en pourcentages. L'homogamie était alors mesurée à la proportion des couples parmi lesquels les conjoints appartenaient aux mêmes groupes (taux d'endogamie absolu). Ces auteurs étaient cependant conscients du fait que cette proportion ne dépendait pas uniquement des proximités et des distances entre groupes sociaux, mais surtout de la répartition des hommes et des femmes entre les différentes catégories. Divers indices visant à contrôler cette influence structurelle étaient donc utilisés. Comme nous le verrons au chapitre 1, des développements méthodologiques ultérieurs ont cependant montré que ces techniques n'étaient pas véritablement adaptés à cet objectif. Une seconde limite de ces travaux tient à la qualité des données disponibles : les échantillons n'étaient pas toujours construits de manière rigoureuse, et les comparaisons entre pays ou dans le temps ne pouvaient souvent reposer que sur des harmonisations grossières. Un acquis solide de cette génération tient cependant à la mise en évidence d'une tendance à l'homogamie au-delà de ce qu'un choix aléatoire du conjoint impliquerait selon une grande variété de caractéristiques et dans de nombreuses sociétés.

Une seconde génération peut être datée de l'ouvrage de Peter M. Bau et Otis D. Duncan *The American Occupational Structure* (1967), dont un chapitre traite de l'homogamie. Les apports de ces deux auteurs sont multiples. Premièrement, ils ont mobilisé des données relativement fines, à partir d'échantillons représentatifs de grande taille. Deuxièmement, après avoir codé les professions des pères des conjoints sur une échelle continue de statut socioéconomique, ils ont étudié conjointement l'homogamie d'éducation et d'origine sociale à l'aide de modèles de régression linéaire. Cette approche leur a permis de montrer que l'homogamie d'origine sociale était due pour un tiers à l'existence de l'homogamie d'éducation – résultat que l'on peut inversement interpréter comme la preuve de l'existence d'une association entre les origines sociales des conjoints au-delà de celle liée à leur similarité en termes de diplôme. Néanmoins, cette génération laissait elle aussi des difficultés non résolues. La méthode de la régression linéaire ne contrôle pas elle non plus parfaitement l'influence de la distribution des individus des deux sexes entre les catégories sur l'homogamie, ou seulement si des hypothèses relativement fortes concernant cette distribution sont respectées. En outre, l'étude des variations temporelles était uniquement

fondée sur la comparaison de cohortes de naissance, ce qui introduit comme nous le verrons au chapitre 1 des risques de biais importants.

Ainsi, une troisième génération est apparue à la fin des années 1970, en mobilisant l'innovation méthodologique représentée par la modélisation log-linéaire, et développant le recours aux *odds ratios*. Ces techniques ont enfin permis de contrôler correctement l'influence de la structure de la population sur l'homogamie, et de se concentrer ainsi sur des questions comme celle des évolutions temporelles de l'homogamie et de l'hypergamie au-delà de ces contraintes structurelles (homogamie et hypergamie relatives)⁷, ou encore la structure des proximités et distances entre groupes. Les travaux de cette génération se sont aussi attachés à décrire et à expliquer les variations entre pays afin de mettre à l'épreuve les théories présentées ci-dessus.

Enfin, une quatrième génération commence dans les années 1990 (Smeenk, 1998 ; Treiman et Ganzeboom, 2000), et vise à étudier les déterminants du choix du conjoint à travers des approches et des questions plus variées grâce à l'acquis méthodologique de la génération précédente et des données d'une qualité toujours croissante. Si la frontière avec la génération précédente n'est pas toujours très nette, on peut notamment rattacher à cette nouvelle génération des analyses comparatives récentes fondée sur des enquêtes internationales, mais aussi des études des déterminants de l'homogamie au niveau individuel. Cette génération a aussi cherché à identifier les mécanismes de l'évolution temporelle de l'homogamie, en portant son attention sur le lien entre massification scolaire et évolution de l'homogamie d'éducation à l'aide de modèles d'analyse des biographies (Smeenk, 1998 ; Blossfeld et Timm, 2003b). D'autres chercheurs ont repris de la deuxième génération le principe d'étudier plusieurs dimensions de l'homogamie simultanément, mais en tirant profit des méthodes élaborées par la troisième génération (Ultee, Dessens et Jansen, 1988 ; Kalmijn, 1994 ; Uunk, 1996). Ces deux dernières approches n'ont cependant pas véritablement été reprises par des travaux ultérieurs.

Cette thèse s'inscrit dans la quatrième vague de travaux, en s'interrogeant à la fois sur les variations dans le temps et dans l'espace du choix du conjoint, sur ses mécanismes au

7 Du point de vue méthodologique, l'ouvrage de Peter Blau et Joseph Schwartz (1984), pourtant très novateur du point de vue théorique, appartient toujours à la deuxième génération, car il ne mobilise pas ces nouvelles techniques pour contrôler les variations de la structure de la population, ce qui aurait permis d'évacuer d'emblée nombre de difficultés.

niveau individuel et ses déterminants au niveau macrosocial. Pour cela, elle mobilise différentes approches généralement retenues par cette génération : la modélisation log-linéaire et log-multiplicative ; les modèles d'analyse biographique (*event-history analysis*) ; la comparaison internationale portant sur un grand nombre de sociétés à l'aide d'enquêtes internationales.

6 Plan de la thèse

Notre étude s'articule en trois parties. Du fait de l'abondance de la littérature internationale sur le sujet et de son caractère fortement technique, nous présentons les travaux existants progressivement, notamment au début de chaque partie, en lien avec la progression de nos analyses. Toutes les analyses portent sur l'ensemble des couples cohabitants, mariés ou non.

La première partie traite des évolutions annuelles de l'homogamie en France entre 1969 et 2011, en termes de diplôme, de classe et d'origine sociales, mais aussi de salaire et de statut d'emploi, à partir des enquêtes Emploi conduites par l'Insee.

Le premier chapitre dresse une revue de la littérature internationale, et particulièrement française, concernant les évolutions temporelles de l'homogamie, présente les principes d'harmonisation de la série de données, la question de la mesure de l'homogamie, puis détaille les évolutions de la structure de la population et celle des couples en termes de diplôme, de classe et d'origine sociales. Il met en évidence une baisse de la proportion de couples appartenant au même groupe social suivant les trois composantes étudiées, ainsi que des transformations majeures de la composition des couples.

Le chapitre 2 reprend les données du précédent pour analyser les évolutions de l'homogamie une fois contrôlées celles de la structure de la population (homogamie relative), d'abord concernant son niveau global, puis en distinguant les différents groupes et plusieurs dimensions de l'espace social. Confirmant et approfondissant les résultats du chapitre 1, il montre que l'homogamie relative a sensiblement diminué selon toutes les composantes ; seule la tendance à l'entre-soi des diplômés des grandes écoles s'est renforcée au cours du temps.

Le chapitre 3 clôt la première partie en étendant l'étude à des dimensions alternatives de la stratification sociale dont on pourrait soupçonner qu'elles aient pris plus d'importance dans les années récentes : le salaire et le statut d'emploi (précarité et chômage) des conjoints. Il apporte une confirmation des résultats précédents, en montrant que l'augmentation de l'activité féminine n'a pas entraîné une plus forte association entre salaires des conjoints, et a même contribué à une diminution des inégalités entre couples. Concernant le statut d'emploi, aucune évolution n'est à relever, excluant l'hypothèse de l'apparition d'un groupe social stable correspondant au « précarité ».

Alors que la première partie se concentrait sur les individus en couple, la deuxième partie étend le champ d'étude à l'ensemble des individus, pour ainsi étudier les déterminants de la mise en couple et du célibat. Cet élargissement est en effet nécessaire pour étudier l'hypergamie féminine, qui est directement liée à la question du célibat différencié selon le sexe.

Le chapitre 4 étudie l'évolution de l'hypergamie de diplôme, de classe et d'origine sociale et des différences entre hommes et femmes dans le taux célibat selon le groupe social en France entre 1969 et 2011, toujours à partir des enquêtes Emploi. Il montre que la position des femmes par rapport à leur conjoint a fortement progressé sur la période, l'écart s'étant même inversé en termes de diplôme. L'hypergamie une fois contrôlée la structure de la population (hypergamie relative) a presque disparu en termes de classe sociale, et s'est là aussi inversée du point de vue du diplôme. Aucune hypergamie selon l'origine sociale n'a pu être observée. Illustrant le lien entre hypergamie et célibat, les différences genrées dans le taux de célibat selon le groupe social se sont largement résorbées.

Le chapitre 5 vise à mieux comprendre les mécanismes à l'origine de la relation entre hypergamie et célibat à travers une étude de la mise en couple fondée sur les enquêtes Biographies et entourage (Ined, 2001) et Étude de l'histoire familiale (Insee-Ined, 1999). Après avoir confirmé, par une étude portant sur les cohortes nées entre 1920 et 1970, les évolutions relevées au chapitre 4, il montre que le surcroît de célibat des diplômées tenait à une forte chute des chances de mise en couple des femmes au-delà de 25 ans ; ce phénomène a disparu parmi les cohortes nées après les années 1950.

Le chapitre 6 prolonge l'étude des mécanismes du choix du conjoint en analysant à partir des deux mêmes enquêtes l'effet de l'âge et de la scolarisation sur l'homogamie et l'hypergamie, en mettant l'accent sur l'influence de la structure du marché conjugal sur les opportunités de rencontre. Il confirme que la scolarisation augmente les chances de former un couple avec un conjoint de même niveau d'éducation (quoique de manière modeste), et que l'hypergamie augmente avec l'âge de mise en couple.

La troisième partie passe d'une comparaison dans le temps à une comparaison dans l'espace, en portant l'attention sur les variations et les déterminants de l'homogamie et de l'hypergamie parmi 64 régions de l'Union européenne. Elle est l'occasion de mettre à l'épreuve les théories développées plus haut dont la validité a été confirmée par les deux parties précédentes dans le cas de la France.

Le chapitre 7 commence par une revue de littérature concernant les analyses comparatives existantes, puis présente la mesure de l'homogamie et de l'hypergamie en termes d'éducation et de groupe socioéconomique à partir de l'Enquête européenne sur les forces de travail (*European Labour Force Survey*) 2011 d'Eurostat. Il met ensuite en évidence de fortes différences dans le niveau aussi bien que dans la structure de l'homogamie entre pays ainsi qu'entre régions d'un même pays. L'hypergamie présente elle aussi des variations notables, et s'inverse même dans plusieurs régions.

Le chapitre 8 se fonde sur les mesures de l'homogamie et de l'hypergamie relatives calculées au chapitre précédent pour analyser les déterminants du choix du conjoint en Europe. Il recourt pour cela à un grand nombre d'indicateurs opérationnalisant les hypothèses retenues tout au long de la thèse, mesurés notamment à partir des vagues 1 à 5 de l'Enquête sociale européenne (*European Social Survey*). Le cadre développé ci-dessus apparaît largement validé, malgré des nuances.

PARTIE I

DIMENSIONS ET ÉVOLUTIONS DE L'HOMOGAMIE EN FRANCE (1969-2011)

CHAPITRE 1

DE LA STRUCTURE DE LA POPULATION À CELLE DES COUPLES : LE DÉCLIN DE L’HOMOGAMIE ABSOLUE DE DIPLÔME, DE CLASSE ET D’ORIGINE SOCIALES

Si les recherches concernant les évolutions de l’homogamie au cours des dernières décennies forment un champ relativement développé au niveau international, les travaux portant sur la France sont restés peu nombreux¹. L’objet de ce chapitre sera donc de dresser une première description de l’homogamie et de ses évolutions entre 1969 et 2011 selon trois dimensions fondamentales de la stratification sociale : le diplôme, la classe sociale et l’origine sociale des conjoints.

Le choix de ces trois dimensions, pour classiques qu’elles soient, est cependant relativement original dans la littérature. En effet, en raison du faible taux d’activité des femmes de l’époque, les premières études portant sur la France (Girard, 1964 ; Desrosières, 1978 ; Deville, 1981 ; Thélot, 1982) retenaient le plus souvent une approche asymétrique, en croisant la catégorie socioprofessionnelle du mari et celle de son beau-père, plutôt que les professions des deux conjoints, ou que celles de leurs pères, comme nous le faisons ici. Suite à des débats importants au cours des années 1980 concernant cette approche, et plus largement à propos de la prise en compte du statut professionnel des femmes (Delphy, 1977 ; de Singly, 1987 ; Vallet, 1986, 2001), les études les plus récentes de l’homogamie ont adopté cette lecture symétrique entre hommes et femmes. Celle-ci présente de plus l’avantage – du moins en ce qui concerne l’étude de l’homogamie de diplôme et de classe sociale – de distinguer clairement choix du conjoint et mobilité sociale intergénérationnelle.

¹ Des éléments de ce chapitre ainsi que du suivant ont fait l’objet d’une publication dans la *Revue française de sociologie* sous le titre « Les évolutions de l’homogamie de diplôme, de classe et d’origine sociales en France (1969-2011) : ouverture d’ensemble, repli des élites » (Bouchet-Valat, 2014).

Les critères retenus ont aussi leur importance. Or, comme on le verra plus bas, en dehors du cas de la France, les études des évolutions temporelles de l'homogamie retenant une autre dimension que le niveau d'éducation des conjoints sont particulièrement rares. C'est aussi le cas des études associant l'examen de plusieurs critères. Ce n'est pourtant qu'en retenant des aspects assez divers de la stratification sociale qu'il est possible de traiter précisément les hypothèses portant sur la potentielle augmentation de l'homogamie suivant les caractéristiques acquises, et la diminution de celle liée aux caractéristiques héritées.

Après une revue de la littérature internationale détaillant les limites des travaux existants sur le cas français, nous présentons en détail les enquêtes Emploi sur lesquelles nous nous appuyons dans l'ensemble de cette partie, ainsi que la méthode d'identification des couples et la définition des caractéristiques des conjoints à partir de ces données. Nous développons ensuite les deux mesures possibles de l'homogamie : l'approche absolue, qui consiste à mesurer la composition des couples en termes de proportions ; puis l'approche relative, mesurant la sur- ou sous-représentation des configurations de couple – en termes d'*odds ratios* – au-delà de ce qu'impose la structure des populations masculine et féminine. La dernière partie de ce chapitre est consacrée à l'étude de l'évolution de la composition des populations masculine et féminine, ainsi qu'à celle du niveau et de la structure de l'homogamie absolue selon le diplôme, la classe et l'origine sociales. Nous mettons en évidence une transformation importante de la composition des couples, et notamment une diminution de la proportion de couples endogames selon les trois dimensions étudiées qui va au-delà de ce que l'évolution de la structure de la population aurait impliqué. Ces premiers résultats indiquent un affaiblissement notable de l'homogamie qui sera analysé plus précisément au chapitre suivant.

1 Des résultats en débat

1.1 La littérature internationale : des conclusions divergentes

Le cadre théorique le plus souvent retenu dans l'étude de l'évolution temporelle de l'homogamie dérive des théories de la modernisation présentées en introduction. Néanmoins, les travaux existants, qui se sont largement concentrés sur l'homogamie d'éducation, ont fourni jusqu'ici des résultats relativement peu cohérents, même entre pays

très similaires. Ce constat pourrait indiquer qu'un cadre aussi large que les théories de la modernisation n'est pas approprié pour expliquer à lui seul l'évolution de l'homogamie d'éducation (Blossfeld, 2009).

En effet, plusieurs travaux portant sur les États-Unis et le Canada ont mis en évidence un renforcement de l'homogamie d'éducation depuis plusieurs décennies (Mare, 1991 ; Kalmijn, 1991 ; Qian et Preston, 1993 ; Pencavel, 1998 ; Schwartz et Mare, 2005 ; Hou et Myles, 2008). À l'inverse, les travaux portant sur l'Europe de l'Ouest utilisant des méthodes similaires ont observé des évolutions divergentes. Un affaiblissement de l'homogamie d'éducation a été mesuré en Grande-Bretagne, en Suède et en Norvège (Halpin et Chan, 2003 ; Birkelund et Heldal, 2003 ; Henz et Jonsson, 2003 notes 9 et 12), ainsi qu'en Finlande, en France et aux Pays-Bas (Ultee et Luijkx, 1990, p. 135). Une baisse concernant les moins diplômés et une hausse pour les plus diplômés a été signalée aux Pays-Bas et en Espagne (Hendrickx, 1998 ; Esteve et Cortina, 2006). En Irlande, une augmentation suivie d'une stabilisation a été observée (Halpin et Chan, 2003) ; en Allemagne a été mesurée une augmentation pour les moins diplômés, mais pas de tendance claire pour les autres groupes (Grave et Schmidt, 2012). Une augmentation globale de l'homogamie d'éducation a même été signalée en Flandres, aux Pays-Bas, en Italie et en Espagne (Ultee et Luijkx, 1990 ; Hendrickx, 1998 ; Bernardi, 2003 ; Esteve et Cortina, 2006). Enfin, des pays européens de l'ancien bloc soviétique (Hongrie, Pologne, République tchèque et Slovaquie), ont de leur côté fait l'objet de plusieurs études qui observent soit une stabilité, soit une baisse suivie d'un certain renforcement de l'homogamie d'éducation après la chute du Mur (Katrňák, Kreidl et Fónadová, 2006 ; Katrňák, 2008).

Dans d'autres parties du monde, des études ont aussi observé diverses tendances. Ainsi, une stabilité de l'homogamie relative d'éducation a été relevée en Australie (Jones, 1987). Son affaiblissement a été mesuré dans dix pays asiatiques (Smits et Park, 2009). Au contraire, c'est une augmentation de l'homogamie des plus diplômés qui a été observée au Brésil et au Mexique (Esteve et McCaa, 2007), de même qu'une augmentation de l'homogamie aux deux extrêmes de l'échelle des diplômes en Argentine (Rodríguez, 2014).

Ce tableau complexe est encore brouillé par la fragilité du point de vue méthodologique de certains des résultats repris ici : certaines des apparentes divergences, parfois entre pays ayant connu des trajectoires similaires, que nous avons résumées ici, pourraient bien

découler de ce type de différences (Blossfeld, 2009). Nous développerons plus bas certaines de ces difficultés méthodologiques à propos du cas de la France. Soulignons seulement ici la question essentielle de l'inclusion ou non des couples cohabitants non mariés, qui pourrait perturber les comparaisons dans le temps et entre pays. En effet, plusieurs études ont mis en évidence une moindre homogamie d'éducation parmi les couples cohabitants que parmi les couples mariés (Blackwell et Lichter, 2000 pour les États-Unis ; Hamplová et Le Bourdais, 2008 pour le Canada), différence qui s'atténue lorsque la cohabitation hors mariage se banalise (Hamplová, 2009 pour l'Europe ; Esteve, McCaa et López, 2013 pour l'Amérique latine)². Les écarts importants entre pays concernant la proportion de couples mariés peuvent conduire les études qui se fondent sur cette seule forme d'union à surestimer l'homogamie dans les pays dans lesquels la cohabitation hors mariage est la plus courante. Au sein d'un même pays, le développement temporel des cohabitations peut quant à lui entraîner, en se fondant sur les seuls couples mariés, une surestimation de l'homogamie parmi les cohortes ou dans les années les plus récentes. De même, les variations avec l'âge du taux de cohabitation peuvent biaiser les études qui se fondent sur ce critère pour mesurer l'évolution temporelle de l'homogamie. Ce biais a pour effet paradoxal d'augmenter l'homogamie apparente dans les pays, les cohortes ou les années d'enquête connaissant les taux de cohabitation les plus élevés, qui sont potentiellement ceux dans lesquels elle serait la plus faible du fait du lien déjà mentionné entre cohabitation et intensité de l'homogamie. Afin d'assurer la comparabilité des résultats, il importe donc de retenir l'ensemble des couples, quel que soit leur statut légal.

L'évolution temporelle de l'homogamie de classe sociale n'a pas bénéficié du même niveau de recherches théoriques et empiriques ; néanmoins, le débat autour de la modernisation peut aussi être appliqué à cette composante (Smits, Ultee et Lammers, 1999). Les quelques travaux la concernant ont observé une diminution de cette dernière dans huit pays d'Europe de l'Ouest (Smits, Ultee et Lammers, 1999), ou encore une baisse statistiquement non significative aux Pays-Bas (Verbakel, Luijkx et de Graaf, 2008).

Le cas de l'homogamie d'origine sociale est un peu différent puisqu'il s'agit d'une caractéristique héritée : nous pouvons raisonnablement nous attendre à ce que les identités

2 D'autres études ont observé une différence de même sens mais non significative (López-Ruiz, Esteve et Cabré, 2009 pour l'Amérique latine), voire aucune différence significative (Blackwell et Lichter, 2000). Nos propres analyses (non reprises ici) font apparaître une homogamie relative (*cf. infra*) légèrement moindre parmi les couples cohabitants que parmi les couples mariés en France.

acquises prédominent sur celles qui sont assignées à la naissance (Blau et Duncan, 1967, chap. 10 ; Kalmijn, 1991 ; Forsé et Chauvel, 1995 ; Mäenpää, 2014), en particulier du fait que le contrôle parental sur le choix du conjoint a perdu de sa force. De plus, étant donné que la mobilité sociale (aussi bien absolue que relative) a augmenté et que le lien entre origine sociale et diplôme obtenu s'est affaibli en France au cours des cinquante dernières années (Vallet, 2004), nous devrions nous attendre à une intensification des relations entre groupes sociaux, du fait du développement de « l'entrecroisement des cercles sociaux » (Blau et Schwartz, 1984) : les conjoints ne peuvent pas être homogames sous tous les aspects à la fois, ce qui implique nécessairement un affaiblissement de l'homogamie sur une partie de ses composantes. On peut penser qu'en cas d'impossibilité à respecter les deux appartenances, les caractéristiques acquises primeront sur les caractéristiques héritées. De fait, l'homogamie d'origine sociale diminue selon les quelques travaux qui l'ont étudiée, à la fois en Europe et aux États-Unis (Kalmijn, 1991 ; Forsé et Chauvel, 1995 ; Uunk, 1996, chap. 3 ; Verbakel, Luijkx et de Graaf, 2008).

Finalement, il est important de noter qu'à l'inverse de l'éducation, qui est largement présente dans les discours portant sur la société, la conscience d'appartenir à une classe sociale a baissé en France, en particulier pour la classe ouvrière, alors que la part des « classes moyennes » – une non-classe – a augmenté au sein des déclarations (Michelat et Simon, 2004, chap. II.1 ; Amossé et Chardon, 2006). Aussi, si les dimensions subjectives ont une importance (quel que soit le sens de la relation entre les deux phénomènes), nous pourrions nous attendre à ce que l'homogamie de classe sociale et d'origine sociale diminue plus fortement que celle de diplôme, rejoignant ainsi une hypothèse de Michel Forsé et Louis Chauvel (1995, p. 141) concernant la prédominance des stratifications à caractère culturel sur celles à caractère socioprofessionnel.

1.2 Les études sur la France : résultats et limites

Les études sociologiques portant sur l'homogamie en France remontent aux années 1960, lorsque Alain Girard (1964), analysant les résultats de son enquête *Le Choix du conjoint*, a souligné l'importance persistante de l'homogamie, en contradiction avec une liberté croissante dans le choix du conjoint chez les jeunes, ainsi qu'avec le développement de la mobilité, des moyens de communication et de l'urbanisation. En recourant à un indice

composite incluant l'origine géographique, la nationalité, la profession, la profession du père, le niveau d'éducation et l'appartenance religieuse, il observait que le niveau global de ressemblance entre conjoints était plus faible chez les jeunes couples que chez les couples âgés, différence qu'il interprétait comme le signe d'une diminution de l'homogamie au cours du temps.

À l'échelle plus réduite d'un village de l'Eure, mais sur une période bien plus longue, Martine Segalen et Albert Jacquard (1971) ont observé, à partir des registres d'état civil, une diminution importante de l'homogamie socio-professionnelle entre 1803 et 1962.

La question de la mesure de l'homogamie a de nouveau été abordée dans les années 1980, d'abord par Claude Thélot (1982, chap. 8), qui a observé une baisse de la proportion de couples dans lesquels la profession du mari et celle de son beau-père sont identiques entre 1953 et 1977. Dans le contexte d'une transformation majeure des modèles familiaux et d'une participation croissante des femmes au marché du travail, ce sont les professions des deux conjoints qui ont retenu l'attention d'autres travaux. Ainsi, Pierre-Alain Audirac (1982) a cherché à mesurer si les couples cohabitants étaient moins homogames que les couples mariés en utilisant de simples pourcentages, et a conclu que, si l'hétérogamie semblait légèrement plus forte parmi les couples cohabitants, il était trop tôt pour tirer des conclusions définitives. Louis-André Vallet (1986), introduisant les modèles log-linéaires et utilisant les données de quatre recensements, a observé une diminution de l'homogamie socioprofessionnelle parmi les couples composés de deux actifs, en termes absolus aussi bien que relatifs.

Les travaux plus récents cherchant à mesurer l'évolution de l'homogamie au cours du temps (Forsé et Chauvel, 1995 ; Goux et Maurin, 2003 ; Vanderschelden, 2006a) présentent tous deux faiblesses communes. Premièrement, à l'instar de l'étude d'Alain Girard, ils se sont fondés chacun sur une seule enquête, utilisant l'âge des interrogés comme une mesure de l'évolution temporelle. Cette méthode est fragile puisqu'elle ne tient pas compte de la légère différence de taux de divorce qui a été observée en France entre couples homogames et couples hétérogames (Commaille, 1974 ; Kellerhals *et al.*, 1985 ; voir aussi Mäenpää et Jalovaara, 2014 pour la Finlande). Le biais ainsi introduit tendrait, le cas échéant, à augmenter artificiellement l'homogamie mesurée chez les couples les plus âgés : puisqu'ils ont été exposés au risque d'une séparation plus longtemps, ils apparaissent dans l'enquête

comme plus homogames qu'ils ne l'étaient lorsqu'ils se sont formés, les couples hétérogames étant sous-représentés dans l'échantillon du fait de plus fréquentes séparations (Kalmijn, 1991, p. 500). Nous pourrions alors conclure à une stabilité de l'homogamie alors qu'elle aurait en réalité augmenté. Aux âges élevés, ce biais est aggravé par l'existence d'une mortalité différentielle, qui amène à surestimer la part des classes supérieures, dont l'espérance de vie est la plus élevée, particulièrement parmi les générations les plus anciennes (Blanpain, 2011).

Un problème plus général concernant l'utilisation de données transversales avec lesquelles l'âge sert à approcher le changement temporel tient à ce que les individus tendent, en vieillissant, à modifier – généralement vers le haut – le diplôme (Baudelot, 1989) et l'origine sociale (Thélot, 1983) qu'ils déclarent, ce qui pourrait avoir toutes sortes de conséquences inattendues sur la mesure de l'homogamie – même si sans doute dans une mesure limitée.

Le second problème tient au fait que ces travaux ont retenu des classifications extrêmement simplifiées. L'éducation a été appréhendée en quatre catégories (aucun diplôme, primaire, secondaire, supérieur) : sans distinctions plus fines, il est impossible d'observer de potentiels changements au sommet de la distribution ou entre les filières générale et professionnelle. Des auteurs ont aussi remarqué que des évolutions divergentes pouvaient être observées selon le niveau d'agrégation retenu (Wong, 2003 pour Taïwan ; Hou et Myles, 2008 pour les États-Unis). De même, la profession a été mesurée à l'aide du niveau 1 des PCS en six catégories, ou à l'aide d'une classification encore plus simple, qui ne permet en particulier pas de distinguer travailleurs qualifiés, non qualifiés (Amossé et Chardon, 2006) et agricoles, ou encore les indépendants selon le nombre de salariés qu'ils emploient.

Ces limites mises en évidence, résumons les méthodes et résultats de ces études. Michel Forsé et Louis Chauvel (1995) ont utilisé une mesure dérivée de l'*odds ratio*, l'indice de diagonalité nette, et les données de l'enquête Emploi 1989. Ils ont identifié une baisse de l'homogamie d'origine sociale, et une stabilité de l'homogamie de diplôme en termes relatifs. Au-delà des problèmes déjà soulignés ci-dessus, une limite de cette étude tient à ce que l'indice de diagonalité nette ne pondère pas les *odds ratios* de la table d'homogamie en fonction des effectifs qu'ils représentent dans la population en moyenne sur la période. En

utilisant une telle pondération, nous avons pu observer un léger affaiblissement de l'homogamie de diplôme dans la même enquête (Bouchet-Valat, 2011).

Plus récemment, Dominique Goux et Éric Maurin (2003) ont conclu, en utilisant l'enquête Formation et qualification professionnelle 2003, et en comparant les taux d'endogamie d'éducation observés à ceux qui seraient attendus sous l'hypothèse de choix aléatoire du conjoint, que la réduction des taux observés entre les cohortes 1934-38 et 1959-63 avait été plus faible que la réduction des taux attendus – résultant donc en un renforcement de l'endogamie relative. Cette méthode inhabituelle (adoptée dans l'ensemble de l'ouvrage, Blossfeld et Timm, 2003) fournit des valeurs assez irrégulières au cours du temps, sans intervalles de confiance ; aucune explication de l'absence de recours aux *odds ratios* n'était fournie. Finalement, seuls les couples mariés étaient considérés, en dépit du développement rapide des unions libres en France au cours des dernières décennies.

Enfin, en ayant recours à des modèles log-linéaires et log-multiplicatifs plus sophistiqués, Mélanie Vanderschelden (2006a) a étudié l'homogamie d'éducation et l'homogamie socioprofessionnelle, en utilisant l'enquête Étude de l'histoire familiale 1999. Elle a mesuré une baisse irrégulière de l'homogamie d'éducation mais aucun changement du niveau de l'homogamie socioprofessionnelle (tous deux en termes relatifs). Malheureusement, les données souffraient de problèmes supplémentaires. Concernant l'homogamie d'éducation, à cause de la structure de l'enquête, seules les premières unions qui n'avaient pas connu de séparation avant l'enquête étaient prises en compte. Concernant l'homogamie socioprofessionnelle, la dernière union de l'interrogé était utilisée, même si le couple était séparé au moment de l'enquête ; la profession utilisée correspondait au moment de l'enquête pour l'interrogé, et à une date non déterminée pour le (dernier) conjoint.

Tenant une synthèse de ces travaux, nous pouvons remarquer qu'un seul d'entre eux (Goux et Maurin, 2003) conclut, sans mesure explicite de l'ampleur du changement, à une augmentation de l'homogamie d'éducation en termes relatifs. La seule étude (Vallet, 1986) qui a eu recours à des enquêtes conduites à des dates différentes a mesuré une diminution en termes à la fois relatifs et absolus de l'homogamie socioprofessionnelle.

2 La construction d'une série harmonisée des enquêtes Emploi

À la lumière de la discussion de la littérature existante que nous venons de réaliser, il apparaît que l'amélioration de nos connaissances concernant l'évolution de l'homogamie au cours du temps exige le recours à plusieurs enquêtes réalisées à des dates différentes. Plusieurs enquêtes de la statistique publique française peuvent être envisagées.

2.1 Les sources disponibles

La source la plus directement reliée à la mise en couple, l'état civil, est fortement limitée par le fait qu'elle n'enregistre que les mariages, et que les caractéristiques des conjoints sont observées uniquement à la date de cet événement. En outre, si le bulletin de mariage renseigne la profession des époux et de leurs parents, aucune information n'est disponible concernant les diplômés.

Autre source de référence, les recensements de la population, déjà utilisés pour l'étude de l'homogamie socioprofessionnelle par L.-A. Vallet (1986), offrent l'avantage d'effectifs très importants et d'une répétition à intervalles réguliers (1968, 1975, 1982, 1990, 1999, et en continu depuis 2004). Par ailleurs, les enquêtes Famille leur sont adossées, fournissant potentiellement des informations complémentaires tenant aux parcours conjugaux. Mais ces enquêtes portaient exclusivement sur les femmes jusqu'à l'extension aux hommes en 1999, et ne contiennent jusqu'en 1990 d'informations que concernant les conjoints mariés. Seules les données du recensement proprement dit, par l'appariement des bulletins individuels remplis par les conjoints cohabitants mariés ou non, pourraient donc être utilisées.

Néanmoins, sans apport propre des enquêtes Famille, les données du recensement ne présentent, par rapport à celles d'enquêtes par sondage, que l'avantage des effectifs élevés. En contrepartie, les données proviennent d'un questionnaire auto-administré relativement fruste (en particulier en ce qui concerne le diplôme), et rempli sans l'assistance d'un enquêteur, ce qui peut entraîner une moindre précision des déclarations et des recodages – notamment en ce qui concerne la question ouverte portant sur la profession. Étant donné la sensibilité de la mesure de l'homogamie à la qualité du codage, il semble plus intéressant de recourir à des enquêtes de moindre ampleur, mais dont la qualité peut être présumée plus élevée.

La série d'enquêtes Formation et qualification professionnelle (FQP) correspond à ce profil. Elle couvre les années 1964, 1970, 1977, 1985, 1993 et 2003 avec un dispositif d'enquête resté très stable dans le temps, raison pour laquelle elle a classiquement été utilisée pour l'étude de l'évolution de la mobilité sociale intergénérationnelle (Thélot, 1982 ; Merllié et Prévot, 1997 ; Vallet, 1999). Son orientation vers les questions de qualification et d'emploi assure une grande qualité et un grand détail du codage du diplôme et de la profession des conjoints, qui répondent individuellement, mais dont les bulletins peuvent être appariés au sein d'un même ménage (du moins pour les enquêtes les plus récentes). Finalement, les échantillons, d'une taille variant entre 18 000 et 40 000 individus selon les années, apparaissent suffisants pour une étude relativement fine de l'homogamie et de ses évolutions. En revanche, la dernière enquête en date (2003) est déjà relativement ancienne, la prochaine (2015) étant actuellement en cours de passation.

2.2 Les enquêtes Emploi

Nous avons donc finalement choisi de retenir les enquêtes Emploi de l'Insee³. Comme les enquêtes FQP, cette série garantit une grande précision des informations concernant les conjoints, collectées par un enquêteur dans le cadre d'un questionnaire approfondi. Commencée en 1950, ses fichiers sont disponibles sous forme informatique à partir de 1962, et l'appariement des conjoints est possible depuis 1969⁴. Sa périodicité annuelle, puis trimestrielle depuis 2003, nous permet de couvrir les années récentes avec des échantillons toujours supérieurs à 40 000 individus.

La principale difficulté associée à l'utilisation de cette enquête tient au fait que son dispositif, qui n'avait pas été conçu pour assurer des comparaisons de long terme, a été remanié à quatre occasions (Goux, 2003), exigeant un travail de mise en cohérence. Néanmoins, grâce à la disponibilité de données annuelles, nous pouvons mobiliser 42 enquêtes réparties en cinq séries complètement homogènes (1969-74⁵, 1975-81, 1982-89,

3 Nous tenons à remercier l'ADISP (Centre Maurice Halbwachs) qui a mis les fichiers de données à notre disposition *via* le réseau Quetelet.

4 L'enquête de 1968 suit théoriquement le même modèle que les suivantes, mais le fichier informatique ne contient pas le diplôme des individus. Nous l'avons donc exclue de notre série.

5 Du fait de l'absence de variables fondamentales et de la mauvaise qualité des données dans l'enquête de 1973, cette année a dû être retirée de la série.

1990-2002, 2003-2011), et vérifier leur cohérence par l'examen des potentielles ruptures de série – possibilité qui n'est pas ouverte lors de l'utilisation d'enquêtes ponctuelles.

Chaque enquête porte sur un échantillon de 40 000 à 75 000 « logements ordinaires ». Cette définition exclut les collectivités (internats, hôpitaux, maisons de retraite, prisons...) et les habitations mobiles ; néanmoins, les individus résidant de manière seulement occasionnelle dans un logement ordinaire sont rattachés à celui-ci. Dans chaque logement, un questionnaire est rempli pour chaque personne de plus de 15 ans au cours d'un entretien. En cas d'absence, un membre du ménage répond à la place d'un autre (proxy), ce qui peut affecter la qualité des renseignements ; cependant, la présence d'un agent enquêteur garantit un niveau minimal de qualité, probablement plus élevé que celui des recensements auto-administrés. Des travaux ont montré que cette particularité du dispositif n'a pas d'impact négatif important sur la déclaration des diplômes (Vallet et Selz, 2007). Le taux de réponse à l'enquête est toujours supérieur à 85 %, et généralement supérieur à 90 %.

Le dispositif d'enquête a connu plusieurs remaniements au fil des ans. De 1969 à 2002, l'enquête était annuelle, et l'échantillon renouvelé par tiers : chaque logement demeurait donc dans l'échantillon pendant trois ans, que ses occupants aient changé ou non. Cependant, une seconde enquête, menée à l'automne, a été expérimentée entre 1975 et 1981 ; par souci de simplicité, les effectifs étant suffisants, nous ne l'avons pas mobilisée. Depuis 2003, l'enquête, dite « en continu », est trimestrielle, avec un renouvellement par sixièmes : chaque logement demeure ainsi 18 mois dans l'échantillon. Afin d'éliminer les doubles comptes, nous ne retenons que la première interrogation de chaque ménage⁶.

Le plan de sondage de l'enquête Emploi, aréolaire à deux degrés, est relativement complexe : l'échantillon est constitué de grappes de logements tirées à partir du recensement de la population ; il est utilisé pour toutes les enquêtes d'une série, avant d'être progressivement remplacé au fur et à mesure des rotations⁷. Les communes et les agglomérations sont tout d'abord divisées selon les séries en 48 à 126 strates, constituées par le croisement de la région (entre 8 et 21) et de la tranche d'unité urbaine (5 ou 6). Des

6 Entre 1975 et 1981, les échantillons de l'enquête menée au printemps et de celle menée à l'automne sont disjoints (Goux, 2003, p. 52), si bien que retenir les seules premières interrogations ne pose aucune difficulté particulière dans cette série.

7 Excepté en 2003, où l'enquête utilise l'échantillon entièrement renouvelé de l'enquête pilote de 2002, fondé sur le recensement de 1999. En 2009, le nouvel échantillon a été tiré à partir de la base de la taxe d'habitation 2006, et les strates ont été définies à partir d'un nombre de critères plus important.

secteurs ou groupes d'aires⁸ de 120 à 240 logements proches sont ensuite tirés de manière aléatoire dans chaque strate à proportion de leur nombre de logements⁹. Ces secteurs sont eux-mêmes divisés en 3 à 6 aires (selon les années) d'environ 20 à 40 logements (non nécessairement contiguës, bien que proches, dans certaines séries ; depuis 2003, un secteur peut contenir plus de 6 aires, auquel cas un tirage aléatoire est effectué). Chaque aire est assignée à un sous-échantillon qui demeurera trois ans (puis 18 mois à partir de 2003) dans l'enquête : une année donnée, chaque secteur est donc représenté dans l'échantillon par une seule de ses aires.

Cet échantillon principal, renouvelé après chaque recensement de la population, est complété chaque année par inclusion des logements neufs construits dans les aires enquêtées. Avant 2003, un échantillon spécial est constitué à partir des logements neufs construits sur un permis regroupant plus de dix logements, qui fonctionne sur le même modèle que l'échantillon principal.

Du point de vue de l'exploitation statistique, le plan de sondage de l'enquête Emploi est en fait relativement proche d'un tirage aléatoire simple en ce qui concerne les valeurs des estimateurs. Le redressement de la non-réponse et des fluctuations d'échantillonnage, permis par les pondérations calculées par calage sur marges, ne modifie pas de manière fondamentale les résultats. Ils améliorent cependant légèrement la cohérence des séries au cours du temps, notamment au moment de l'introduction de l'enquête en continu en 2003, qui semble avoir exigé quelques années de rodage. Nous les utilisons donc systématiquement dans les analyses.

Mais c'est surtout du point de vue de l'incertitude associée aux estimateurs que le plan de sondage complexe gagne à être pris en compte. D'une part, la stratification en régions et tranches d'unité urbaines à partir du recensement implique qu'une observation de l'enquête Emploi contient plus d'informations qu'une observation issue d'un sondage aléatoire simple. Mais d'autre part, le regroupement des logements en secteurs ou aires implique une homogénéité plus forte entre ces logements (corrélation intra-grappe) que dans un

8 La terminologie a varié selon les séries d'enquête. Les anciennes séries parlaient simplement d'aires, l'unité la plus fine étant appelée sous-aire ; les noms des variables reflètent encore cet usage d'origine.

9 En pratique, pour éviter d'avoir à découper l'ensemble du territoire en aires, il a été procédé à un tirage de segments de taille plus importante (canton ou grande ville), puis de sous-segments (commune rurale ou quartier pour les villes) à l'intérieur, finalement divisés de manière exhaustive en secteurs. Dans la mesure où un seul secteur est finalement tiré dans chaque segment, le plan de sondage demeure à deux degrés, l'unité primaire d'échantillonnage étant indifféremment le segment, le sous-segment ou le secteur.

échantillon aléatoire simple, impliquant une information moins riche. C'est le second effet qui l'emporte sur le premier. L'ampleur de la perte d'information peut être évaluée en comparant la précision des estimateurs de quantités simples, comme des proportions, à celle que l'on obtiendrait dans le cadre d'un échantillon de même taille tiré par sondage aléatoire simple. Le rapport entre les variances de ces deux estimateurs, appelé effet du plan de sondage (*design effect*), varie généralement entre 3 et 5, ce qui ne semble pas négligeable et pourrait nous conduire à considérer comme statistiquement significatifs des effets qui ne résultent que du caractère aléatoire du tirage. Aussi, nous avons identifié les strates et les secteurs dans chaque série d'enquête, de manière à ce que les intervalles de confiance et les indicateurs permettant l'inférence statistique tiennent toujours compte du plan de sondage. Le paquet R *survey* (Lumley, 2014) a été utilisé pour le calcul de ces quantités.

Dans ce qui suit, nous présentons la construction des variables de diplôme, de classe sociale et de classe sociale d'origine des conjoints, avant d'explicitier la méthode d'identification des couples et le champ retenu.

2.3 La définition des variables

2.3.1 Le diplôme

L'éducation des individus est définie comme le plus haut diplôme obtenu en formation initiale ou continue. Avant d'aborder les détails concernant sa mesure et sa cohérence à travers les séries d'enquête, soulignons que nous nous intéressons ici à un diplôme validé, et pas seulement à un niveau d'éducation. Nous mesurons donc non seulement des facteurs culturels et des effets liés à la fréquentation du système scolaire, mais aussi des titres scolaires qui ont une valeur institutionnalisée en termes notamment de carrière et de statut social, mais aussi du point de vue conjugal (Weber, [1922] 1968, partie 2, ch. XI, sec. 13 ; Bourdieu, 1989). En ce sens, l'homogamie de diplôme offre une mesure de l'importance des groupes de statut dans une société – des groupes dont la constitution était, selon Goblot (1925, chap. VI), la seule fonction du Baccalauréat.

Nous avons retenu la classification la plus détaillée (en 11 ou 12 catégories, cf. tableau 1.1). qui puisse être reconstituée pour les 42 années d'enquête. Elle correspond de manière relativement directe à la classification internationale CASMIN (Brauns et Steinmann, 1999). Ce choix d'une classification assez détaillée est en cohérence avec l'objectif de distinguer des

variations à différents niveaux de l'échelle des diplômes et sur différentes dimensions, en plus d'une tendance globale, et d'éviter les problèmes liés à l'utilisation de classifications trop agrégées abordés plus haut.

Derrière d'importantes variations dans les noms des variables, les codes et les intitulés, la nomenclature de diplôme utilisée par l'Insee est restée relativement stable, même si des regroupements et des divisions ont été introduits au cours du temps. La reconstruction d'une nomenclature cohérente sur toutes les séries exige cependant de repartir des variables renseignant séparément le diplôme général, technique ou professionnel, et supérieur, de manière par exemple à distinguer les titulaires d'un CAP/BEP conjointement avec un BEPC, distinction qui a été abandonnée en 2003, reflétant la perte de visibilité de ce dernier diplôme dans le nouveau fonctionnement du système d'enseignement.

Quelques changements de définition ont cependant pu être identifiés grâce à l'observation des ruptures de série. Entre 1982 et 2002, nous avons ramené les titulaires d'un Certificat de fin d'études normales, d'un Certificat d'aptitude pédagogique ou d'un Diplôme de maître d'éducation physique au niveau Baccalauréat, plutôt qu'au niveau du premier cycle du supérieur, par cohérence avec les séries précédentes. Après 2003, ces diplômes semblent complètement ignorés par l'enquête, ce qui revient *de facto* à adopter la correction que nous venons de décrire, dans la mesure où la validation d'un cursus à l'École normale s'accompagnait du passage du Baccalauréat général (possédé par 80 % des instituteurs entre 2003 et 2011). De même, dans la série 2003, les titulaires d'un Certificat d'éducation professionnelle sont classés par l'enquête au niveau BEPC, alors que dans les autres séries ce diplôme était ignoré. Nous avons donc rétabli l'ancienne définition.

Tableau 1.1 : La classification de diplôme retenue

Intitulé	Code CASMIN	Détail
Aucun	1a	Aucun diplôme
CEP	1b	Certificat d'études primaires, Diplôme de fin d'études obligatoires
CAP/BEP	1c/2a	CAP/BEP, Certificat de formation professionnelle pour adultes
CAP/BEP +BEPC	2a+b	CAP/BEP + BEPC
BEPC	2b	BEPC, BEPS, Brevet élémentaire
Bac gén	2c_gen	Baccalauréat général, Capacité en droit, Diplôme de fin d'études normales, Certificat d'aptitude pédagogique ou Diplôme de maître d'éducation physique
Bac pro/tech	2c_voc	Baccalauréat technologique ou professionnel, Brevet professionnel, Brevet de maîtrise, Brevet d'enseignement (agricole, commercial, hôtelier, industriel, social)
1 ^{er} cycle pro	3a_voc	BTS, DUT, Diplôme paramédical et social, autre diplôme de technicien niveau Bac+2
1 ^{er} cycle gén	3a_gen	Propédeutique, DEUG, DUEL, DUES, autre diplôme général de niveau Bac+2
2 ^e cycle*	3b_low	Licence, Maîtrise, concours du professorat
3 ^e cycle*	3b_high	Master, DEA, DESS, Doctorat
Grande école	3b_school	Grande école, école d'ingénieur ou de commerce, diplôme d'avocat, de notaire, d'expert-comptable

* La distinction entre diplômes universitaires du 2^e cycle (Licence, Maîtrise, concours du professorat) et du 3^e cycle (DESS, DEA, Doctorat), qui semble d'un grand intérêt pour l'étude de la période récente, n'a été introduite qu'en 1982. Nous fusionnons donc ces deux niveaux pour l'étude des évolutions temporelles, mais les distinguons dans les analyses transversales portant sur la fin de la période.

La question de la cohérence des catégories au cours du temps se pose avec une acuité particulière concernant les grandes écoles, du fait des multiples définitions que peut recouvrir ce terme, mais aussi des résultats que nous obtenons, qui font de ce groupe une exception. Il importe donc de s'assurer de la comparabilité de ce groupe entre les séries. Une première garantie est offerte par l'absence de rupture dans l'évolution des effectifs de ce groupe au cours du temps (Figure 1.2, p. 94). Concernant la définition elle-même de la catégorie, les dictionnaires des codes fournissent des indications plus ou moins détaillées

selon les séries, mais toujours cohérentes (Tableau 1.2) : jusqu'en 2003 étaient distingués d'une part un groupe constitué des écoles les plus prestigieuses (dont on possède la liste entre 1982 et 2002), et d'autre part l'ensemble des écoles scientifiques, commerciales ou littéraires, auxquels étaient rattachés les diplômes d'ingénieurs obtenus hors de ce cadre, ainsi que ceux d'avocat, de notaire et d'expert comptable. En fusionnant ces deux catégories, la définition que nous retenons est donc extensive, mais déjà suffisamment précise pour une étude en population générale¹⁰ : ce groupe représente entre 0 et 6 % des individus selon le sexe et l'année d'enquête.

En 2003, l'introduction d'un nouveau mode de collecte, plus précis, s'est accompagnée de l'abandon de la distinction entre les plus prestigieuses des grandes écoles et les autres, au profit d'une tripartition entre écoles d'ingénieurs, écoles de commerce, et autres diplômes au-delà de Bac+3 (hors université). En effet, depuis le passage à l'enquête en continu, les cartes codes proposant une liste fermée de catégories ont été remplacées par une saisie en clair des intitulés des formations, fournissant une classification très détaillée qui est finalement agrégée de manière automatisée avant mise à disposition des fichiers¹¹. On voit cependant que le champ de la catégorie « grandes écoles » est resté le même.

10 Dans leur étude portant sur les inégalités d'accès aux grandes écoles, Valérie Albouy et Thomas Wanecq (2003) ont mis à profit cette double définition, extensive et restrictive, des grandes écoles.

11 Nous tenons à remercier Daniel Martinelli, responsable du pôle Diplômes et spécialités de formation de l'Insee, pour nous avoir fourni ces précisions.

Tableau 1.2 : La description des grandes écoles dans les dictionnaires des codes

Série	Variable, modalité et intitulé
1968-1974	EG 8 : Diplôme de sortie d'une grande école scientifique, commerciale ou littéraire (voir la liste dans les Instructions aux enquêteurs [liste non incluse dans la documentation]) EP 8 : Diplôme d'ingénieur attribué par une école ne figurant pas dans la liste d'instruction aux enquêteurs.
1975-1981	EG 8 : Diplôme de sortie d'une grande école EP 8 : Diplôme d'ingénieur
1982-1989	DIES 7 : Diplôme d'une grande école ou école d'ingénieur (non cité ci-dessous) Diplôme d'ingénieur du CNAM, diplôme d'études comptables supérieures (DECS), diplôme d'avocat (CAPA), d'expert-comptable DIES 9 : Diplôme de l'une des grandes écoles suivantes : Normale Sup, Polytechnique, Centrale, Mines, Ponts, ENGREF, Télécom, Sup-Aéro, Génie Maritime, ENSAE première division, ENA, Sciences Po Paris, HEC, INA, Navale, École de l'Air, Saint-Cyr, ESSEC, École de la Magistrature
1990-2002	DIES 48 : Diplôme d'une grande école ou école d'ingénieur (non cité ci-dessous) y compris d'études comptables supérieures (DECS), d'avocat (CAPA), d'expert-comptable, de 2 ^e cycle de notariat DIES 49 : Diplôme de l'une des grandes écoles suivantes : Centrale, École de l'Air, École de la Magistrature, ESSEC, ENA, ENGREF, ENSAE première division, Génie maritime, HEC, INA (« Agro »), Mines, Navale, Normale Supérieure, Polytechnique, Ponts, Saint-Cyr, Sciences Po Paris, Sup-Aéro, Télécom
2003-2011	DIP 12 : Écoles de niveau licence et au-delà, dont : – DIPSUP 55 : Autres diplômes supérieurs (niveau Bac+3 et plus) – DIPSUP 63 : Écoles d'ingénieur – DIPSUP 64 : Écoles de commerce

2.3.2 La classe EGP

La reconstruction d'une variable harmonisée de classe sociale sur les cinq séries d'enquêtes a exigé un travail bien plus laborieux. La rupture entre la nomenclature des catégories socioprofessionnelles (CSP) et celle des professions et catégories socioprofessionnelles (PCS) introduite en 1982 rendait nécessaire un recodage d'au moins une partie des enquêtes. Reprenant le codage minutieux réalisé par Louis-André Vallet sur les enquêtes Formation et qualification professionnelle, nous utilisons ici la classification Erikson-Goldthorpe-Portocarero (EGP). Outre la justification pratique tenant à l'existence d'un programme de codage, le recours à cette classification permet la comparaison directe de nos résultats avec d'autres pays, la nomenclature EGP étant le standard de fait des études de la stratification sociale. Plus fondamentalement, par son niveau de détail intermédiaire (11 catégories) entre le niveau 1 (6 catégories) et le niveau 2 (24 catégories) des PCS, elle représente un bon compromis entre finesse d'analyse et complexité des modèles et des résultats. En particulier, la distinction entre employés et ouvriers qualifiés et non qualifiés paraît de première importance. Si nous utiliserons donc principalement la classe EGP dans nos analyses, nous présenterons aussi ponctuellement des résultats obtenus à l'aide de la PCS agrégée ou détaillée lorsque cela semblera utile, soit pour les résumer sous une forme plus classique pour l'étude de la société française, soit pour les mettre à l'épreuve et les préciser.

Encadré 1.1 : La nomenclature de classe Erikson-Goldthorpe-Portocarero

Le schéma de classe Erikson-Goldthorpe-Portocarero (EGP; Erikson, Goldthorpe et Portocarero, 1979) est une nomenclature internationale, résultat d'une élaboration progressive à partir de travaux réalisés par John Goldthorpe dans les années 1970 en Grande-Bretagne, poursuivis par un effort d'harmonisation de données nationales réalisé dans le cadre du projet CASMIN (*Comparative Analysis of Social Mobility in Industrial Nations*; voir Erikson et Goldthorpe, 1992). Elle est depuis devenue la classification de référence des études comparatives, notamment dans le domaine de la mobilité sociale. Visant à l'origine à regrouper des individus à la situation de marché (revenus, stabilité de l'emploi, perspectives de carrière) et de travail similaires, elle a peu à peu fait l'objet d'une systématisation théorique autour du concept de relation d'emploi. Celui-ci oppose

une classe de service (*service class*, cf. Bidou-Zachariassen, 2000) disposant d'une grande autonomie dans son activité et d'une carrière programmée sur le long-terme, à des travailleurs au « contrat » (*labour contract*) dont l'archétype est le paiement à la pièce (ou, en termes contemporains, le *zero-hour contract*). La notion de *service class* a notamment été reprise en France par Paul Bouffartigue (2001) sous le nom de « salariat de confiance » pour décrire le statut de cadre, ce qui témoigne de sa validité dans des contextes nationaux variés.

La classification EGP a été présentée successivement en plusieurs versions, de 7 à 11 positions. La version la plus détaillée se décline comme suit :

- I : Professions libérales et intellectuelles, cadres administratifs, techniques et commerciaux (niveau supérieur) ; directeurs et administrateurs d'entreprises, gros indépendants
- II : Professions libérales et intellectuelles, cadres administratifs, techniques et commerciaux (niveau inférieur) ; techniciens de niveau supérieur ; contremaîtres dirigeant des employés
- IIIa : Employés de niveau supérieur (administrations et affaires).
- IIIb : Employés de niveau inférieur (commerces et services).
- IVa : Petits indépendants avec salariés (artisans, commerçants, chefs de petites entreprises...).
- IVb : Petits 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 semi- et non-qualifiés hors agriculture.
- VIIb : Ouvriers agricoles et autres ouvriers du secteur primaire.

Si différents que soient les soubassements théoriques de la nomenclature EGP et des PCS, on peut noter une similarité certaine entre les deux classifications, et en particulier une correspondance presque parfaite entre plusieurs catégories : cadres et professions intellectuelles supérieures (I, qui comprend cependant les chefs de très grandes entreprises), agriculteurs exploitants (IVc), artisans, commerçants et chefs d'entreprise (IVb, IVa avec une division à un seuil de salariés différent des PCS, la frange supérieure étant classée en I), ouvriers qualifiés (VI) et non qualifiés (VIIa), ouvriers agricoles (VIIb). Les divergences principales concernent les professions intermédiaires et les employés (catégories majoritairement féminines), qui se trouvent répartis entre quatre classes EGP différentes (II, IIIa, IIIb et V) ; il n'est sans doute pas inutile de rappeler que la définition du groupe des employés au sein des PCS a d'ailleurs fait l'objet de débats (Burnod et Chenu, 2001 ; Amossé et Chardon, 2006).

L'identification de la classe EGP en 11 catégories est réalisée à partir d'une combinaison de la CSP ou de la PCS à deux chiffres, du statut de salarié ou d'indépendant, du nombre de salariés, de la qualification, et dans certains cas de la profession détaillée et du secteur d'activité (le détail du calcul de cette variable est disponible en annexe A, p. 517). Lorsque les inactifs sont pris dans le champ d'étude, les informations relatives à leur dernière profession sont utilisées. Chez les femmes notamment, le sens de ce classement mérite réflexion, la moitié des femmes entre 35 et 54 ans étant inactives encore en 1975 (Vallet, 1986, p. 662) : la définition de la classe sociale des femmes à partir de leur profession pose problème pour le début de la période étudiée, et ce, que l'on se limite aux couples bi-actifs ou que l'on se fonde sur la dernière profession occupée par les inactifs. En effet, il est évident que celle-ci n'a pas la même signification pour une inactive qui n'a travaillé que quelques années avant son mariage, et pour une femme qui poursuit une carrière.

Pour résoudre cette difficulté, nous étudierons par la suite d'une part les couples dans lesquels les deux conjoints sont actifs (occupés ou au chômage), et d'autre part l'ensemble des couples dans lesquels les deux conjoints ont déjà occupé un emploi, en utilisant le dernier emploi pour les chômeurs et les inactifs. Nous serons ainsi en mesure d'évaluer si la dernière profession constitue un bon indicateur de la classe sociale des inactifs, et si l'entrée massive des femmes sur le marché du travail sur la période étudiée a un effet sur

l'homogamie. Ces observations pourront aussi être confrontées aux résultats obtenus à partir de la classe sociale d'origine, qui offre une autre manière de définir la position sociale des individus.

La question proche du classement des aides familiaux (en fait, des aides familiales) doit elle aussi faire l'objet d'une réflexion (Amossé, 2004). Composée de conjointes d'indépendants (très peu d'hommes sont déclarés sous ce statut), cette catégorie a une signification assez délicate pour la mesure de l'homogamie, puisqu'elle est la seule dans laquelle le statut d'un conjoint dépend de celui de l'autre. Classer les aides familiales dans le même groupe que leur conjoint n'est cependant pas aussi tautologique qu'il n'y paraît. D'une part, cette définition reflète une certaine réalité sociale, soit le fait que la conjointe d'un indépendant, quelle que soit la nature exacte de l'activité qu'elle pratique au sein de l'entreprise, ne peut être considérée comme n'importe quel salarié du point de vue du statut d'emploi, qui est l'un des critères fondamentaux de toute classification socioprofessionnelle. D'autre part, cette situation n'est le fait que d'une partie des indépendants : l'endogamie n'est pas postulée par définition¹², et elle s'avère plus ou moins importante selon les groupes. Il reste qu'en faisant prévaloir le statut d'emploi sur le contenu concret de l'activité, le classement des aides familiales laisse échapper des distinctions qui permettraient de définir de manière plus fine la position sociale des conjointes d'indépendants (par leur profession détaillée, ainsi que le proposait Amossé, 2004) ; néanmoins, ce choix paraît en totale cohérence avec le regroupement des indépendants dans une catégorie spécifique, sans égards pour le type d'activité qu'ils pratiquent. En tout état de cause, les enquêtes Emploi ne permettent pas, même dans la série la plus récente, d'expérimenter une approche différente.

Les recodages réalisés pour identifier la classe EGP des individus suivent deux logiques différentes : il s'agit d'une part d'appliquer les critères de la classification EGP (statut, nombre de salariés, qualification...) pour séparer les individus classés dans une même CSP ou PCS ; et d'autre part, de traiter séparément certaines professions détaillées qui ont changé de catégorie selon les séries de manière à les rendre plus cohérentes. Concernant particulièrement ce dernier point, la disponibilité de données annuelles s'est révélée un

12 L'assimilation mécanique du statut social de la femme à celui de son mari a fait l'objet de vives critiques de la part des féministes (Delphy, 1977 ; Vallet, 2001). Ce n'est pas à cette opération que nous nous livrons ici : elle reviendrait à postuler *a priori* une homogamie de classe sociale parfaite, supprimant du même coup l'objet de notre étude.

atout précieux, permettant d'évaluer la cohérence des recodages effectués entre séries : des ruptures qui seraient passées inaperçues avec des enquêtes ponctuelles ont pu être repérées, menant à des ajustements. C'est par exemple le cas concernant les femmes de ménage et les agents du nettoyage, que la nomenclature des PCS considère comme ouvriers s'ils travaillent pour des entreprises, mais employés s'ils travaillent dans le public et dans les secteurs de la santé, de l'enseignement, de l'hôtellerie, de la restauration ou des services domestiques, alors que la nomenclature des CSP et les Code des métiers de 1968 et 1975 tenaient seulement compte du type d'employeur (secteur public, secteur privé ou particuliers). Ainsi, la moitié des membres de cette catégorie considérés jusque-là comme ouvriers ont été reclassés comme employés en 1982 (Chenu, 1990, p. 29) ; cette rupture de série a pu être corrigée en tenant compte du secteur d'activité.

Contraint par la qualité des informations fournies par les anciennes séries, nous avons généralement privilégié la cohérence du recodage à sa précision, en particulier lorsque les améliorations ne concernaient que la dernière série d'enquête. Dans cette logique, le nombre d'employés des aides familiaux n'est utilisé dans aucune série, étant donné que cette information n'est disponible qu'à partir de 2003 : son introduction aurait créé une trop forte rupture en déplaçant de nombreuses femmes (les plus fréquemment déclarées comme aides familiales) de la classe IVb à la classe IVa. La distinction entre ces deux catégories est donc sujette à caution chez ces dernières. En revanche, le nombre de salariés de l'entreprise de l'emploi antérieur, qui apparaît en 1990, est utilisé à partir de cette enquête pour classer les chômeurs et inactifs anciennement indépendants, étant donné la date relativement précoce de son introduction, et le caractère plus limité de cet ajustement.

Au total, la continuité des différentes séries apparaît très bonne chez les hommes (Figure 1.6, p. 105), mais un peu moins chez les femmes (même si elle reste tout à fait correcte), ce qui reflète la définition généralement plus floue des emplois féminins, ainsi que les tentatives de redéfinition récente dont ils ont fait l'objet (Chenu, 1990 ; Amossé, 2004). Une grande partie des ruptures de séries proviennent en fait de modifications du codage des PCS, notamment en 2003 : elles sont aussi clairement visibles en utilisant la PCS à deux chiffres, et même au niveau des professions détaillées.

En dépit du fait qu'entre 1975 et 1981, la CSP à deux chiffres correspondant à la dernière profession des inactifs n'est pas disponible, et a dû être imputée à partir des catégories à

deux chiffres les plus représentées dans les séries précédente et suivante, la continuité s'avère aussi bonne lorsque les inactifs sont pris en compte – nous en verrons plusieurs illustrations plus loin.

2.3.3 La classe EGP d'origine

La classe sociale d'origine ne peut être identifiée que depuis 1982. La construction de cette variable est fondée sur la profession du père de l'interrogé (ou mère, ou tuteur en cas d'absence du père à cette époque) au moment où il ou elle a cessé de fréquenter régulièrement l'école ou l'université. Les mêmes règles de recodage sont utilisées que pour la classe sociale des conjoints, mais de manière simplifiée du fait du moindre détail des informations. En effet, la profession détaillée n'est pas renseignée, et le nombre de salariés (pour les indépendants) ne l'est qu'à partir de 1990 : de manière à assurer la cohérence des séries, cette information a donc été ignorée, et les catégories IVa et IVb fusionnées. Pour plus de détails, le lecteur pourra se référer à l'annexe A, p. 517).

2.4 Le choix des limites d'âge et la définition des couples

Nous retenons une définition large des couples, incluant tous les couples hétérosexuels¹³ cohabitants, qu'ils soient mariés ou non. Cette définition s'impose dans le contexte français où la cohabitation s'est fortement développée sur la période considérée, et a même pour certains couples remplacé le mariage (Villeneuve-Gokalp, 1990 ; Toulemon, 1996 ; Prioux, 2005) : nous avons donc des raisons de penser que s'en tenir aux couples mariés introduirait un biais temporel assez net, du fait de la moindre intensité de l'homogamie parmi les couples non mariés déjà mentionnée dans la revue de littérature. Notre définition n'est cependant pas la plus large possible : les couples non cohabitants en sont exclus. On estime que ces unions représentent un peu moins de 10 % des couples, principalement parmi les moins de 35 ans, et cette part semble relativement stable au cours du temps (Beaujouan, Régnier-Loilier et Villeneuve-Gokalp, 2009 ; Buisson et Lapinte, 2013).

Étudier les unions cohabitantes à un instant donné permet de représenter de manière fidèle les nouveaux modes de vie en couple, sans discriminer entre les plus durables et les

13 Depuis 1990, il est possible de se déclarer en couple avec un conjoint de même sexe. Ces couples ont été exclus du champ pour permettre l'étude de l'hypergamie féminine, ainsi que pour des raisons de cohérence avec les anciennes séries. Leur nombre (en augmentation, allant jusqu'à 800 dans l'enquête 2011) semble cependant laisser ouverte la possibilité d'une étude spécifique de cette population.

plus éphémères, mais sans dépendre non plus d'un jugement des individus concernant la signification du fait de vivre en couple, qui a pu évoluer sur la période étudiée. S'en tenir au critère objectif de la cohabitation, qui était le seul retenu en dehors du mariage par la série des enquêtes Famille jusqu'en 2011, permet de maintenir une certaine stabilité dans la définition du couple, malgré les nombreuses transformations que celui-ci a connues sur la période.

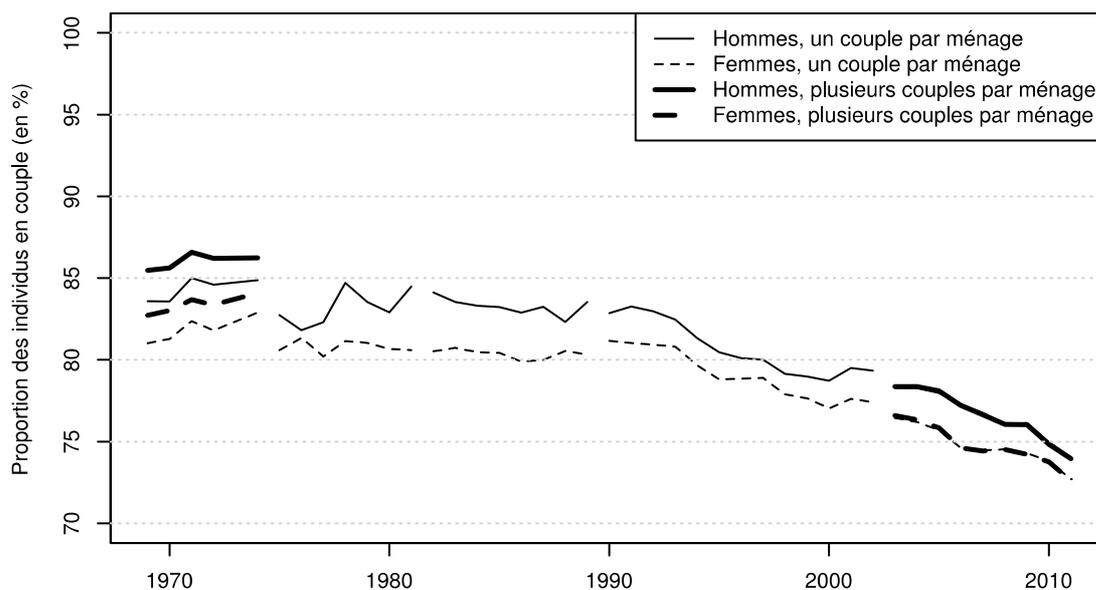
Plus précisément, du fait de la structure des données, nous ne pouvons identifier dans chaque ménage que le couple formé par la personne de référence¹⁴ et son ou sa conjoint(e). Néanmoins, la première et la dernière séries d'enquête nous permettent de distinguer plusieurs couples par ménage, et ainsi d'évaluer si cette restriction est à l'origine d'une distorsion importante. Entre 1969 et 1974, en effet, l'enquête distingue à l'intérieur d'un ménage plusieurs familles (jusqu'à trois), chacune ayant un chef désigné ; une variable indique à quel « chef de famille » un individu est rattaché, et une autre le lien entre l'individu et le « chef de famille ». Depuis 2003, une variable indique simplement quel est l'identifiant du conjoint de l'individu au sein du ménage, permettant d'identifier les couples sans restrictions. Il apparaît que la restriction du champ à un seul couple par ménage n'a qu'un effet très limité (Figure 1.1, différence entre courbes fines et épaisses) : en 1969, seuls 3 % des couples cohabitent avec un autre au sein d'un même ménage ; presque aucun couple n'est dans ce cas en 2011.

Nous appliquons finalement des limites d'âge, en ne retenant que les couples dans lesquels l'un des deux conjoints est âgé de 30 à 59 ans. Cette limitation du champ permet d'éviter une sur-représentation des groupes sociaux se mettant en couple le plus tôt (Robert-Bobée et Mazuy, 2005), ainsi que celles ayant l'espérance de vie la plus élevée (Blanpain, 2011). En outre, la déclaration de la profession des retraités (comme celle de l'ensemble des inactifs) n'est pas aussi précise que celle des actifs dans certaines séries (*cf. infra*). L'application des limites d'âge à l'un ou l'autre des conjoints, plutôt qu'aux deux, vise à éviter d'exclure plus fréquemment les couples à l'écart d'âge élevé, dont la proportion a

14 Dite « chef de ménage » dans les enquêtes anciennes, la personne de référence 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 (Courson, 1982). Le caractère sexiste de cette définition (Amossé et De Peretti, 2011) ne pose aucun problème ici, puisque nous décomposons toujours le ménage en un homme et une femme, plutôt que de chercher à le décrire par la seule personne de référence.

diminué avec le temps (Vanderschelden, 2006b), ce qui serait susceptible d'introduire un léger biais si ces couples sont aussi atypiques du point de vue de l'homogamie.

Figure 1.1 : Évolution de la proportion d'hommes et de femmes en couple



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : individus âgés de 30 à 59 ans, et couples dans lesquels au moins l'un des conjoints remplit cette condition

À l'intérieur de ces bornes d'âge, on peut constater sur la figure 1.1 l'homogénéité relativement élevée des séries, à l'exception de la première, qui paraît sur-estimer systématiquement la proportion de personnes en couple – mais pas cependant de manière rédhibitoire. Cette sur-estimation s'explique probablement par une moins bonne couverture des ménages d'une seule personne que dans les séries suivantes.

En cohérence avec l'augmentation de l'âge de première mise en couple, et le léger accroissement de la proportion de personnes n'ayant jamais vécu en couple (Prioux, 2003), la part de personnes vivant en couple une année donnée a baissé, d'abord lentement, puis à un rythme plus rapide depuis 1990. Elle passe ainsi de près de 85 % chez les hommes et un

peu plus de 80 % chez les femmes à un peu moins de 75 % chez les deux sexes, avec un taux toujours légèrement plus faible chez les femmes. Ces chiffres sont très proches de ceux calculés à partir du recensement de la population (Daguet et Niel, 2010).

La déclinaison par âge et année d'enquête des proportions de personnes en couple (Tableau 1.3) confirme que la tranche d'âge 30-59 ans présente les taux les plus élevés, même si le taux de personnes en couple a diminué à tous les âges au fil des enquêtes. Le taux de vie en couple parmi les 20-29 ans est nettement moins élevé que celui des 30-39 ans pour toutes les années d'enquête. C'est aussi le cas des femmes de 60 à 69 ans, du fait de la mortalité plus précoce chez les hommes.

Tableau 1.3 : Proportion des individus vivant en couple cohabitant (marié ou non) par âge et décennie d'enquête (en %)

Âge	Hommes				Femmes			
	1970-79	1980-89	1990-99	2000-09	1970-79	1980-89	1990-99	2000-09
20-29	44	39	33	33	60	54	48	48
30-39	82	82	77	74	85	82	80	77
40-49	84	85	83	78	84	83	81	76
50-59	86	84	84	80	76	77	78	73
60-69	82	84	83	82	58	63	67	68

Source : enquêtes Emploi 1969-2011 (Insee)

Champ : individus âgés de 20 à 69 ans

Le tableau 1.4 dresse le panorama des cohortes présentes dans notre échantillon pour chaque décennie d'enquête après l'application des limites d'âge. Au total, nos données permettent de décrire les cohortes nées des années 1910 aux années 1970, avec un léger aperçu concernant celles nées dans les années 1900 et 1980. Les transformations de l'homogamie que nous observons résultent à la fois du nouveau comportement des cohortes les plus récentes, et des séparations et potentielles remises en couple au sein des cohortes déjà présentes dans l'échantillon. Ce dernier phénomène ne pose aucun problème pour notre étude étant donné que nous sommes à même de décrire, pour chaque année d'enquête, l'ensemble des couples existant dans la société, quelle que soit l'histoire conjugale des deux partenaires. Si les couples formés à la suite d'une séparation sont différents des autres, cet effet sera pris en compte dans l'évolution globale – sans qu'il soit possible ici de l'analyser

séparément. Pour conclure, notons qu'en ce qui concerne l'homogamie de classe sociale – et dans une faible mesure celle de diplôme –, les caractéristiques des individus se modifient aussi au cours de la vie, évolution qui peut dépendre en partie des caractéristiques du conjoint (Bozon et Héran, 2006, chap. 5) : de même que pour les séparations, nous ne nous intéressons ici qu'à la ressemblance des conjoints chaque année d'enquête, sans tenir compte de leurs trajectoires.

Tableau 1.4 : Répartition des individus par année d'enquête et cohorte de naissance (% en ligne)

Enquête	Cohorte de naissance									Total
	1900-09	1910-19	1920-29	1930-39	1940-49	1950-59	1960-69	1970-79	1980-89	
1969	5	22	34	33	5					100
1970-79	2	12	32	31	21	2				100
1980-89		1	15	27	31	24	2			100
1990-99			1	14	28	34	21	1		100
2000-09				1	17	32	31	18	1	100
2010-11					4	28	32	28	7	100

L'échantillon final, après application des limites d'âge, est composé de 445 000 couples. Après exclusion des valeurs manquantes, restent 441 000 couples (99 %) pour le diplôme, 400 000 (90 %) pour la classe sociale, 260 000 pour la classe sociale des actifs (58 %), et 293 000 (93 % des couples depuis 1982) pour la classe sociale d'origine.

3 De l'homogamie absolue à l'homogamie relative

À la suite de nombreux travaux, nous entendons aller au-delà d'une lecture binaire de l'homogamie opposant couples homogames et couples hétérogames. En effet de nombreuses configurations de couple sont représentées sur l'ensemble de l'espace social : dans les tables d'homogamie présentées plus bas, qui croisent la catégorie d'appartenance de l'homme (en lignes) avec celle de la femme (en colonnes), plus d'une centaine de configurations différentes sont distinguées, autorisant ainsi des mesures fines des distances sociales. Nous avons donc recours à deux approches différentes de l'homogamie.

D'une part, nous désignons le degré le plus élevé d'homogamie, dans lequel les conjoints appartiennent au même groupe, sous le nom d'*endogamie*¹⁵. La tendance à l'endogamie des différentes catégories fournit une mesure du degré de fermeture de ces groupes, qui peut aussi être considérée comme liée à la force de l'identité de groupe. Le taux absolu d'endogamie dépend clairement de la nomenclature retenue ; en particulier, plus les catégories utilisées sont fines, et plus l'endogamie absolue apparaît faible (ce qui n'est pas nécessairement le cas de l'endogamie relative, que nous présenterons plus bas).

D'autre part, nous considérons l'homogamie et l'hétérogamie comme des situations relatives sur un continuum couvrant l'ensemble de l'espace social. Les configurations de couple correspondant à des degrés divers d'homogamie ou d'hétérogamie offrent une information sur la distance sociale entre les groupes auxquels appartiennent les conjoints : celle-ci peut être appréhendée comme l'inverse de la propension des personnes issues de ces groupes à former des couples ensemble. Bien que naturellement toujours sensible à la nomenclature utilisée, ce concept est nettement plus robuste aux changements de définition des catégories puisqu'il ne dépend pas exclusivement du choix des frontières des groupes, répondant ainsi à une difficulté soulignée aussi bien par A. Girard (1964, sect. « Les déterminations sociales du choix ») que par M. Bozon et F. Héran (1988, p. 143).

Cette double conception de l'homogamie trouve aussi de solides justifications empiriques. En effet, les configurations endogames obéissent très nettement à des règles différentes du reste de la table d'homogamie, un phénomène mis en évidence à l'origine par les études de la mobilité sociale à travers le fait que des modèles traitant ces configurations de manière spécifique décrivent de manière plus adéquate les données (*cf.* par exemple Goodman, 1965 ; Duncan, 1979 ; Kalmijn, 1991).

15 Formé à partir des racines grecques ἔνδον (à l'intérieur) et γάμος (mariage), ce terme nous paraît bien refléter la distinction sémantique avec le terme homogamie, dont le préfixe ὁμός (semblable) indique une ressemblance moins stricte entre conjoints. Cet usage du terme endogamie, qui semble trouver son origine au XIX^e siècle chez l'ethnologue anglais John MacLennan, se retrouve chez différents auteurs (Berent, 1954 ; Van Leeuwen et Maas, 2005), même si la littérature francophone contemporaine tend à ne l'appliquer qu'aux ethniques, religieux, géographiques ou de parenté (Bozon et Héran, 1987a, p. 338 ; Widmer, 1993 ; Santelli et Collet, 2011). À l'inverse, la tradition bourdieusienne parlait d'endogamie aussi souvent que d'homogamie à propos des classes sociales (Bourdieu, 1974, 1979, 1984 ; Desrosières, 1978).

3.1 La mesure absolue de l'homogamie : intérêt et limites

En interaction partielle avec la définition de l'homogamie retenue, deux manières très différentes mais complémentaires de mesurer l'homogamie peuvent être utilisées. La mesure la plus élémentaire est le taux *absolu* d'endogamie, c'est-à-dire la proportion de couples dans lesquels les conjoints appartiennent à la même catégorie ; ce principe peut être étendu à la proportion d'une configuration de couple quelconque parmi l'ensemble des couples. Cette approche est valide en tant que description des couples qui peuvent effectivement être observés dans la société à un moment donné. Elle constitue la manière correcte d'analyser l'homogamie en se concentrant sur ses conséquences, par exemple sur les contacts entre groupes sociaux, ou encore en termes d'inégalités entre ménages. C'est cette approche que nous mobiliserons principalement dans ce chapitre, afin de décrire précisément la composition des couples et son évolution.

Cependant, malgré son intérêt, cette mesure absolue n'est pas un outil adapté à tous les questionnements. En effet, le taux d'endogamie est principalement déterminé par la composition des populations respectives des hommes et des femmes dans la classification utilisée. Ce taux reflète donc la contrainte imposée par la disponibilité de conjoints potentiels issus des différents groupes ; même si les individus choisissaient leurs conjoints complètement au hasard, une grande part des couples seraient endogames par nécessité mathématique. C'est en référence à cette situation de mise en couple aléatoire, dite hypothèse d'indépendance, que sont construits les indices d'homogamie relative¹⁶.

Formellement, sous l'hypothèse d'indépendance, chaque configuration de couple (cellule de la table d'homogamie) serait représentée à hauteur du produit des proportions d'hommes et de femmes appartenant aux catégories considérées dans la population (dites proportions marginales, du fait qu'elles correspondent aux marges de la table d'homogamie). Dans la notation du tableau 1.5, la proportion attendue sous l'hypothèse d'indépendance dans la cellule située à l'intersection de la ligne h et de la colonne f est donc $m_{hf} = P_{h+}P_{+f}$. Ainsi, le taux d'endogamie attendu, qui est la somme des proportions observées dans les cellules de la diagonale principale de la table (en grisé), est égal au produit scalaire des deux vecteurs de

16 Cette hypothèse de mise en couple au hasard est l'équivalent direct de celle d'égalité des chances ou des destins sociaux des analyses de la mobilité sociale intergénérationnelle, ou encore de celle de panmixie en génétique des populations (Segalen et Jacquard, 1971).

proportions marginales : $\sum_{k=1}^K P_{k+}P_{+k}$ avec K le nombre de lignes et de colonnes (par définition égaux dans une table d'homogamie).

Tableau 1.5 : Notation des proportions des effectifs des différentes cellules dans une table d'homogamie

Hommes	Femmes			Ensemble
	Catégorie 1	Catégorie 2	Catégorie 3	
Catégorie 1	P ₁₁	P ₁₂	P ₁₃	P ₁₊
Catégorie 2	P ₂₁	P ₂₂	P ₂₃	P ₂₊
Catégorie 3	P ₃₁	P ₃₂	P ₃₃	P ₃₊
Ensemble	P ₊₁	P ₊₂	P ₊₃	1

Les cellules grisées sont celles correspondant aux configurations endogames (diagonale principale de la table).

Mais la structure de la population ne détermine pas le seul taux d'endogamie attendu en situation de mise en couple aléatoire : elle contraint aussi l'intervalle dans lequel ce taux est autorisé à varier (Forsé et Chauvel, 1995, p. 127-128). Le taux d'endogamie maximal qu'il soit possible d'atteindre est très rarement de 100 % : ce n'est le cas que si les distributions des hommes et des femmes sont strictement égales ; dans le cas contraire, certains individus sont forcés de trouver un conjoint hors de leur groupe. Le taux d'endogamie minimal, de son côté, est plus facilement égal à 0 % : il suffit pour cela qu'aucune catégorie ne rassemble plus de la moitié des hommes et des femmes réunis – auquel cas il serait impossible à l'ensemble des individus de ce groupe de trouver tous des conjoints dans d'autres catégories¹⁷.

Afin de mesurer les implications concrètes de ces propriétés mathématiques, illustrons-les rapidement à partir d'exemples fictifs mais néanmoins évocateurs de situations réelles que nous analyserons plus bas.

Considérons tout d'abord une représentation caricaturale de la situation concernant le début de la période étudiée, munis par souci de simplicité d'une nomenclature en seulement deux catégories : supposons que seuls 10 % des hommes et des femmes soient titulaires d'un

¹⁷ Les formules de ces bornes sont respectivement $Endo_{\max} = \sum_{k=1}^K \min(P_{k+}, P_{+k})$ et $Endo_{\min} = \sum_{k=1}^K \max(P_{k+} + P_{+k} - 1, 0)$.

diplôme, les 90 % restants n'en possédant aucun. Le tableau 1.6 représente alors la table d'homogamie de diplôme que l'on observerait si les individus de cette population fictive se mettaient en couple parfaitement au hasard¹⁸ : le taux d'endogamie serait dans cette situation de 82 % (81 % + 1 %). Le taux minimal d'homogamie qu'il soit possible d'obtenir est très élevé, de 80 %, et le taux maximal de 100 %.

Tableau 1.6 : Table d'homogamie simplifiée dans une population peu diplômée

En %	Femmes		
Hommes	Non diplômée	Diplômée	Ensemble
Non diplômé	81 / 80 / 90	9 / 10 / 0	90
Diplômé	9 / 10 / 0	1 / 0 / 10	10
Ensemble	90	10	100

Le premier chiffre de la cellule correspond au pourcentage d'individus observé sous l'hypothèse de mise en couple au hasard, le second à celui obtenu lorsque l'endogamie est minimale, le troisième lorsque celle-ci est maximale. Les cellules grisées sont celles correspondant aux configurations endogames (diagonale principale de la table).

Au contraire, dans une société ayant connu une certaine massification scolaire, et dont la population se diviserait à parts égales entre diplômés et non diplômés (Tableau 1.7), le taux d'endogamie attendu en situation d'indépendance ne serait plus que de 50 % (25 % + 25 %). Du fait de l'équilibre parfait entre diplômés et non diplômés, le taux d'endogamie peut varier entre 0 et 100 %. De ce fait, un taux d'endogamie de 80 % signifierait une forte tendance à choisir un conjoint de même niveau de diplôme au-delà de ce que la structure de la population implique – alors que dans la situation précédente, ce taux aurait indiqué une tendance exactement inverse. Ces exemples, pour extrêmes qu'ils soient, ne font que radicaliser des tendances que nous observerons dans la réalité ci-dessous. Comment peut s'interpréter la différence entre ces deux situations-types ?

¹⁸ Par simplicité, on suppose ici que l'ensemble de la population vit en couple.

Tableau 1.7 : Table d'homogamie simplifiée dans une population relativement diplômée

En %	Femmes		
Hommes	Non diplômée	Diplômée	Ensemble
Non diplômé	25 / 0 / 50	25 / 50 / 0	50
Diplômé	25 / 50 / 0	25 / 0 / 50	50
Ensemble	50	50	100

Le premier chiffre de la cellule correspond au pourcentage d'individus observé sous l'hypothèse de mise en couple au hasard, le second à celui obtenu lorsque l'endogamie est minimale, le troisième lorsque celle-ci est maximale. Les cellules grisées sont celles correspondant aux configurations endogames (diagonale principale de la table).

Ce qui diffère entre les deux populations-types des tableaux 1.6 et 1.7, c'est la *concentration* de la distribution des diplômes. Dans le premier cas, les individus se répartissent de manière très déséquilibrée entre les catégories (90 % contre 10 %) ; dans le second, ils sont répartis de manière parfaitement égale. Le taux d'endogamie sous l'hypothèse d'indépendance, égal au produit scalaire des proportions marginales, est minimal lorsque les proportions sont toutes égales : $P_{h+} = P_{+f} = 1/K$ pour tout h et f entre 1 et K (la somme de chacune des distributions étant par définition fixée à 1).

Naturellement, la distribution par diplôme d'une population dépend dans une large mesure du choix d'une classification plus ou moins détaillée selon les régions de l'espace social. Des classifications assez différentes peuvent ainsi donner lieu à des évolutions divergentes au cours du temps du taux d'endogamie en situation de mise en couple au hasard, pour peu que l'une accorde plus de détail à une région en expansion numérique (par exemple, les plus diplômés), quand l'autre ferait le choix inverse.

La concentration n'est pas le seul facteur déterminant le taux d'endogamie en situation de mise en couple au hasard : la *dissimilarité* des distributions des hommes et des femmes, nulle dans les deux exemples précédents, a aussi un rôle majeur. Le tableau 1.8 reprend les distributions marginales du tableau 1.6, mais en inversant la composition de la population des hommes, ce qui introduit une asymétrie entre hommes et femmes : ici, les femmes sont diplômées à 90 %, les hommes à seulement 10 %. Les cellules se trouvent simplement permutées, mais le résultat est nettement différent. Du fait du très fort déséquilibre entre effectifs masculins et féminins au sein de chaque groupe, l'endogamie est très difficile à

réaliser. Le taux d'endogamie sous hypothèse d'indépendance est de seulement 18 % (9 % + 9 %), et il ne peut varier qu'entre 0 et 20 %. Une seconde règle concernant la détermination du taux d'endogamie attendu apparaît donc : si les distributions doivent être fortement concentrées pour que ce taux soit élevé, elles doivent aussi être similaires entre hommes et femmes. À l'inverse, un taux d'endogamie attendu faible peut être obtenu soit avec des distributions très similaires entre hommes et femmes, mais réparties de manière équilibrée entre les catégories ; soit avec des distributions très dissimilaires selon le sexe, mais – par conséquent – assez fortement concentrées.

Tableau 1.8 : Table d'homogamie simplifiée dans une population où les femmes sont plus diplômées que les hommes

En % Hommes	Femmes		Ensemble
	Non diplômée	Diplômée	
Non diplômé	9 / 0 / 10	81 / 90 / 80	90
Diplômé	1 / 10 / 0	9 / 0 / 10	10
Ensemble	10	90	100

Le premier chiffre de la cellule correspond au pourcentage d'individus observé sous l'hypothèse de mise en couple au hasard, le second à celui obtenu lorsque l'endogamie est minimale, le troisième lorsque celle-ci est maximale. Les cellules grisées sont celles correspondant aux configurations endogames (diagonale principale de la table).

Ainsi, si l'on prend au sérieux la signification sociologique du taux d'endogamie absolue, qui présente des effets réels sur la stratification sociale en dépit – ou plutôt en raison – de sa forte détermination par la structure de la population, il importe de comprendre si son évolution est due à celle de cette dernière, ou à une modification des déterminants du choix du conjoint *au-delà* des contraintes structurelles. En outre, il paraît intéressant de savoir si l'évolution du taux d'endogamie attendu en situation de mise en couple aléatoire provient d'une variation de la concentration des distributions des hommes et des femmes, ou d'une variation de la dissimilarité entre ces dernières. Cette question peut être étudiée en comparant le taux d'endogamie attendu sous l'hypothèse d'indépendance à celui qui serait obtenu si les distributions observées chez les deux sexes étaient toutes deux égales d'une part à celle des hommes, d'autre part à celle des femmes¹⁹. Chacun de ces taux

¹⁹ Plusieurs indices de concentration et de dissimilarité sont disponibles, et offrent souvent des

constitue un indice de la concentration de la distribution de chaque sexe, alors que leur écart au taux attendu en situation d'indépendance donne une mesure de leur dissimilarité.

Nous avons jusqu'ici implicitement fait l'hypothèse que l'ensemble de la population vivait en couple, et donc que le taux attendu d'endogamie pouvait être calculé à partir de la structure de la population, sans information concernant les couples proprement dits. La réalité est évidemment plus complexe, du fait de l'existence d'une proportion de célibataires variable selon les groupes sociaux, ainsi que d'individus se mettant en couple avec un conjoint en dehors des bornes d'âge retenues. Ainsi, les taux d'endogamie en situation de mise en couple au hasard calculé à partir de la structure des populations des hommes et des femmes que nous présentons plus bas diffèrent de ceux qui peuvent être calculés à partir des marges des tables d'homogamie (qui excluent les célibataires et incluent des conjoints hors des bornes d'âge²⁰), et reposent eux-mêmes sur une approximation des couples qui pouvaient être formés du fait de l'application des limites d'âge. Ces écarts sont généralement négligeables. Ils méritent cependant une attention particulière pour l'étude des asymétries hommes-femmes et de l'hypergamie, les proportions de célibataires selon les groupes sociaux étant très différentes chez les deux sexes (chapitre 4). Une analyse des taux de célibat sera donc réalisée en complément de celle de l'homogamie.

Nous présentons maintenant le concept d'homogamie relative, qui permet de prolonger l'analyse en contrôlant les effets de la structure de la population, avant de passer à l'analyse empirique suivant la démarche que nous venons de développer.

3.2 Le concept d'homogamie relative et sa mesure

On voit donc aux éléments présentés dans la section précédente que le taux d'endogamie, même s'il décrit une réalité sociale, ne constitue en aucun cas un indice de la préférence des individus et des normes sociales pour la similarité ou de la recherche de l'entre-soi, ni même des opportunités de rencontres qu'ont connues les individus. Ces questions essentielles ne peuvent être traitées qu'à partir de mesures qui contrôlent l'effet des contraintes structurelles. Ces mesures *relatives* de l'homogamie, fondées sur les *odds*

interprétations assez naturelles. Néanmoins, leur lien avec le taux d'endogamie n'est pas direct, ce qui implique que leurs indications peuvent parfois diverger ; il nous a pour cela paru préférable de travailler directement avec le taux d'endogamie lui-même.

20 Notons que, si nous avons ici choisi de retenir ces conjoints qui n'appartiennent pas au champ, les exclure aurait posé le même problème, en excluant du même coup leur conjoint qui, lui, en fait partie.

ratios, sont dites indépendantes des marges du fait qu'elles ne dépendent pas de la répartition des hommes et des femmes entre les catégories.

Développé en sociologie par les études de la mobilité sociale à partir des années 1970, l'usage des méthodes fondées sur l'*odds ratio* est devenu standard pour l'étude de l'homogamie dans la littérature française et internationale. Grâce aux nombreuses propriétés désirables qu'ils présentent (Vallet, 2007), ces outils remplacent avantageusement plusieurs indices élaborés pour tenir compte des contraintes structurelles, mais qui, n'étant pas réellement indépendants des marges de la table, ne contrôlent qu'imparfaitement ce facteur (Forsé et Chauvel, 1995) : il s'agit de l'indice d'immobilité de Boudon, celui de mobilité de Yasuda, ou encore des indices dérivés du ratio de mobilité de Glass.

Insistons en particulier sur ce dernier, qui est généralement désigné sous le nom d'indice d'homogamie dans la littérature portant sur ce sujet, et continue parfois à être utilisé. Cet indicateur se calcule individuellement pour chaque cellule de la table, et rapporte les effectifs observés à ceux attendus en situation d'indépendance – soit, dans la notation du tableau 1.5, $P_{hf}/P_{h+}P_{+f}$ pour la cellule située à l'intersection de la ligne h et de la colonne f . Malgré son attrait lié à sa relative simplicité d'interprétation, qui explique sans doute son utilisation persistante dans des travaux récents (Vanderschelden, 2006a, 2006c), cet indice n'est pas indépendant des marges de la table : il convient donc de lui préférer d'autres mesures.

Dans cette logique, nous recourons ici à l'*odds ratio* généralisé (Kaufman et Schervish, 1987), outil particulièrement adapté à la présente étude, et qui semble mériter une diffusion plus large que celle dont il a bénéficié jusqu'à présent. Nous présentons d'abord rapidement le calcul des *odds ratios* d'une table d'homogamie, puis celui de l'*odds ratio* généralisé, qui sera utilisé plus bas ; la question de la mesure générale de l'intensité de l'homogamie relative sera traitée au chapitre suivant, en même temps que les modèles log-linéaires.

Une table de contingence formée de H lignes et F colonnes, permet de calculer $H(H-1)F(F-1)/4$ *odds ratios* différents en choisissant à chaque fois deux lignes et deux colonnes, soit quatre cellules²¹. Chaque *odds ratio* mesure les chances d'un homme de la catégorie h (ici, une ligne) de se mettre en couple avec une femme de la catégorie f (colonne)

21 Une table d'homogamie étant (généralement) carrée, on a ici $H = F = K$. Nous retiendrons cependant, ici et par la suite, le cas le plus général d'une table rectangulaire.

plutôt que de la catégorie f' , par rapport aux chances d'un homme de la catégorie h' de le faire avec une femme de la catégorie f plutôt que de la catégorie f' . La lecture de l'*odds ratio* peut se faire de manière strictement équivalente en partant d'un sexe ou de l'autre, comme on le voit à sa formule :

$$\theta_{hf,h'f'} = \frac{P_{hf}/P_{hf'}}{P_{h'f}/P_{h'f'}} = \frac{P_{hf}/P_{h'f}}{P_{hf'}/P_{h'f'}} = \frac{P_{hf}P_{h'f'}}{P_{h'f}P_{hf'}}$$

Ainsi, les asymétries genrées se mesurent non pas en lisant un *odds ratio* donné dans un autre sens, mais en comparant deux *odds ratios* différents, mais construits de manière symétrique par rapport à la diagonale de la table. Ces *odds ratios* sont obtenus en inversant les catégories des hommes et des femmes, soit h et f d'un côté, h' et f' de l'autre : par exemple, dans le tableau 1.5 (p. 84) $\theta_{11,32}$ et $\theta_{11,23}$.

L'*odds ratio* évolue entre 0 et $+\infty$. Sous l'hypothèse d'indépendance, tous les *odds ratios* de la table sont égaux à 1 ; un *odds ratio* supérieur à 1 indique une association positive entre les catégories h et f (sur-représentation), un *odds ratio* inférieur à 1 une association négative (sous-représentation). La propriété d'indépendance des marges de l'*odds ratio* s'exprime dans le fait qu'il n'est pas affecté par la multiplication des lignes et/ou des colonnes par des valeurs arbitraires. Ceci s'observe aisément dans la dernière expression de l'équation ci-dessus : un coefficient multipliant l'ensemble des cellules d'une ligne ou d'une colonne se trouvant simultanément au numérateur et au dénominateur, est éliminé de la formule de l'*odds ratio*.

Comme on le voit en comparant les deux premières expressions de l'équation, chaque groupe de quatre cellules permet de calculer deux *odds ratios* différents, chacun étant égal à l'inverse de l'autre ; la dernière expression permet un calcul plus direct. Notons que seuls $(H - 1)(F - 1)$ *odds ratios* sont nécessaires pour calculer tous les autres (dénommés pour cela *basic set*, cf. Goodman, 1969) : ceux-ci peuvent notamment correspondre à tous les *odds ratios* impliquant une cellule de référence et chacune des cellules situées sur une ligne et une colonne différente (*spanning cell odds ratios*, cf. Rudas, 1998), ou à tous ceux construits à partir de quatre cellules adjacentes (*odds ratios locaux*).

L'inconvénient majeur de l'*odds ratio* pour l'étude de tables présentant plus de deux lignes et/ou de deux colonnes – comme la plupart des tables d'homogamie – vient du fait que si chaque cellule peut donner lieu à la construction de $(H - 1)(F - 1)$ *odds ratios*, aucun

d'entre eux ne constitue une mesure de la sur- ou sous-représentation de la configuration de couple correspondant à cette cellule par rapport à ce que l'on attendrait en situation de mise en couple au hasard (soit l'équivalent indépendant des marges de l'indice d'homogamie critiqué plus haut).

L'*odds ratio* généralisé, proposé par Robert Kaufman et Paul Schervish (1987) il y a déjà près de trente ans²², nous semble mériter une plus grande place dans l'analyse des tables de contingence précisément parce qu'il remplit cette fonction essentielle, tout en demeurant de calcul et d'interprétation relativement simples. Cet indice est simplement égal à la moyenne géométrique des $(H - 1)(F - 1)$ *odds ratios* qui peuvent être calculés à partir de la cellule considérée. Soit, en notant h sa ligne et f sa colonne :

$$GOR_{hf} = \left(\prod_{h' \neq h} \prod_{f' \neq f} \theta_{hf, h'f'} \right)^{\frac{1}{(H-1)(F-1)}}$$

Il s'interprète simplement comme les chances pour un homme de la catégorie h par rapport à un homme de toute autre catégorie de se mettre en couple avec une femme de la catégorie f plutôt que de toute autre catégorie – et inversement du point de vue des femmes. Comme l'*odds ratio*, cet indice varie entre 0 et $+\infty$. Néanmoins, il n'est pas nécessairement égal à 1 lorsque les effectifs de la configuration de couple considérée sont égaux à ceux attendus en situation d'indépendance (et vice-versa) : ce n'est le cas que si l'ensemble des cellules de la table respectent cette hypothèse ; en effet, l'*odds ratio* généralisé tient compte de l'association présente dans l'ensemble de la table, et pas seulement dans la cellule considérée (Cobalti, 1989). Cette particularité, qui pourrait paraître indésirable, découle en fait mathématiquement de la propriété d'indépendance des marges²³. L'*odds ratio* généralisé revient en fait à rapporter les effectifs observés à ceux qui seraient attendus si, *les autres cellules étant fixées* (c'est-à-dire l'association dans le reste de la table restant constante), il n'existait aucune association entre la catégorie h et la catégorie f (Cobalti, 1989) ; ou encore

22 Le principe lui-même a été introduit encore plus tôt par Leo Goodman (1969) à propos de l'étude de tables de mobilité sociale ; cependant, seules les cellules de la diagonale étaient concernées, l'indice étant alors désigné sous le nom de *intrinsic status inheritance*.

23 En effet, pour que l'*odds ratio* généralisé soit égal à l'unité, il faut et suffit que les effectifs de la cellule puissent être exprimés uniquement à partir de coefficients affectant de manière multiplicative chaque ligne et chaque colonne de la table – coefficients qui ne sont pas nécessairement égaux aux proportions marginales.

aux effectifs moyens que l'on attendrait à partir de ce qui est observé dans les autres cellules. Cette lecture est assez différente de celle de l'indice d'homogamie (ou ratio de mobilité de Glass), qui compare les effectifs observés à ceux qui le seraient si aucune association n'existait dans aucune cellule de la table.

Ce point sera développé au chapitre suivant, dans lequel nous verrons que l'*odds ratio* généralisé est très directement relié aux coefficients d'interaction du modèle log-linéaire saturé. La formulation en termes d'*odds ratios* retenue ici nous semble intéressante pour sa – relative – simplicité, ainsi que pour son utilisation de l'échelle assez familière des *odds ratios*, offrant une lecture en termes de chances relatives – quand la formulation en termes de modèles log-linéaires recourt souvent à une échelle logarithmique fort abstraite. Notons que nous ne recourons ici aux *odds ratios* généralisés que pour fournir une première illustration de la structure des associations entre groupes sociaux : une analyse plus précise sera réalisée au chapitre suivant.

Dans la dernière section de ce chapitre, nous décrivons l'évolution de la structure des populations des hommes et des femmes âgés de 30 à 59 ans de 1969 à 2011 en termes de diplôme, de classe sociale et d'origine sociale. Nous observons que les taux absolus d'endogamie ont clairement baissé pour ces trois composantes ; les taux qui seraient attendus en situation de choix aléatoire du conjoint ont aussi diminué en ce qui concerne le diplôme, mais pas en ce qui concerne la classe et l'origine sociales.

4 De l'évolution de la structure de la population à celle des couples

4.1 Le diplôme : les effets de la massification scolaire

4.1.1 Une transformation majeure de la structure par diplôme de la population

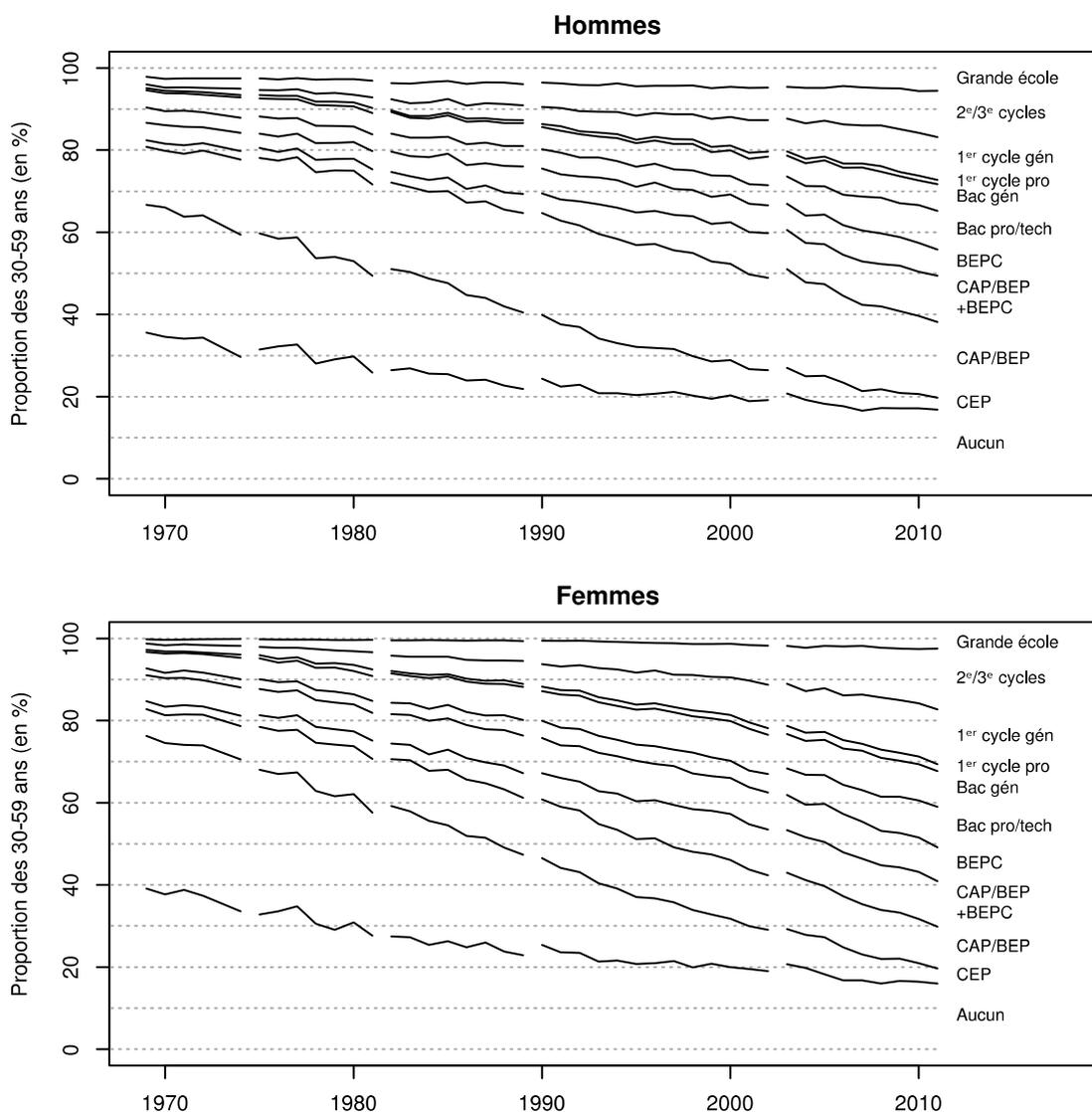
C'est – de manière attendue – du point de vue de l'éducation que la population française s'est le plus fortement transformée entre 1969 et 2011 (Figure 1.2). Les effets successifs de la première (1937-1948) et de la seconde explosion scolaire (après 1968) apparaissent ici lissés par l'inclusion de nombreuses cohortes dans notre échantillon (Chauvel, 1998) : on observe

ainsi une augmentation régulière du niveau moyen de diplôme des 30-59 ans, nés entre 1910 et 1939 au début de la période, et entre 1952 et 1981 à sa fin. L'étude de la population générale plutôt que des cohortes peu après leur sortie du système scolaire offre une vision du passé de ce système, dotée d'une forte inertie. Ainsi, les deux évolutions les plus frappantes sont ici la disparition progressive du groupe des titulaires du certificat d'études primaires (CEP), qui représentaient entre 30 et 40 % de la population étudiée en 1969²⁴, ainsi que la forte réduction de la proportion de non diplômés (passant de 35 % chez les hommes et 40 % chez les femmes à 16 % chez les deux sexes). Cette baisse est la conséquence d'un développement relativement uniforme de tous les autres groupes de diplôme, mais qui a été particulièrement fort pour les diplômes du supérieur, et notamment chez les femmes.

Ces dernières sont ainsi devenues globalement plus diplômées que les hommes (Duru-Bellat, 1990 ; Baudelot et Estabiet, 1992). La figure 1.2 nous permet de dater cette inversion parmi les 30-59 ans : en 1980 déjà, exactement un quart des hommes, et une proportion très légèrement inférieure des femmes possédaient au moins le BEPC ; en 1990, c'était le cas de 35 % des hommes et de 40 % des femmes. La même inversion se retrouve au niveau du Baccalauréat, quoique avec des écarts plus faibles ; elle a été un peu plus tardive pour le supérieur, avec une égalisation autour de l'an 2000. En revanche, il semble intéressant de noter que la proportion de non diplômés est restée pratiquement identique chez les deux sexes tout au long de la période. Le dépassement du niveau d'éducation des hommes par les femmes s'est donc réalisé par le haut, *via* le développement de l'ensemble des diplômes au-dessus du certificat d'études au détriment de cette dernière catégorie, mais sans affecter le bas de la distribution.

24 Rappelons que ce titre scolaire a progressivement disparu entre 1959 (extension de la scolarité obligatoire jusqu'à 16 ans), 1972 (restriction aux seuls adultes) et 1989 (suppression totale).

Figure 1.2 : Évolution de la structure par diplôme des 30-59 ans (proportions cumulées)



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : individus âgés de 30 à 59 ans

Lecture : En 1969, 76 % des femmes âgées de 30 à 59 ans ne possédaient aucun diplôme ou le seul Certificat d'études primaires.

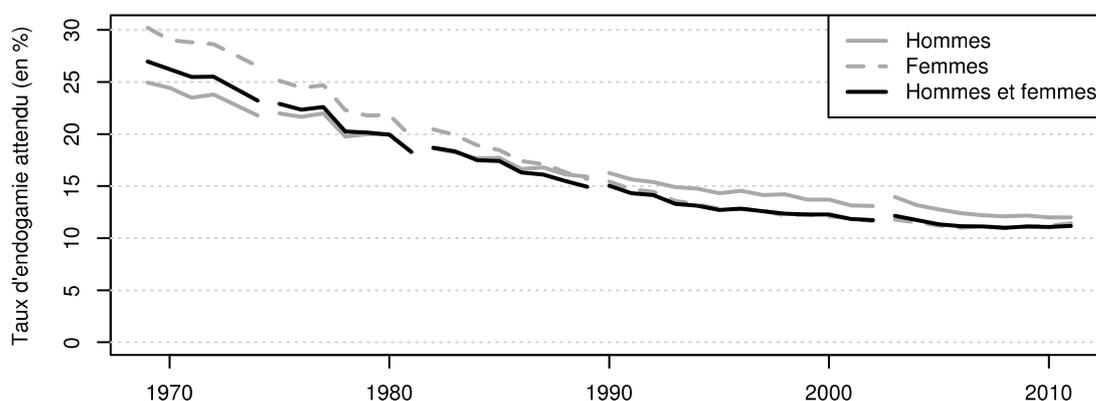
4.1.2 Une nette diminution du taux d'endogamie de diplôme

Cette transformation radicale de la structure par diplôme de la population a nécessairement eu des effets massifs sur le taux d'endogamie. Le taux d'endogamie attendu en situation de mise en couple au hasard a nettement baissé, passant de 27 à 11 % entre 1969 et 2011 (Figure 1.3, courbe noire). Cette évolution s'explique directement par l'effondrement

de la concentration de la distribution par diplôme des hommes et des femmes, due à la forte baisse des effectifs des deux groupes des moins diplômés, les plus représentés au début de la période. En effet, les taux d'endogamie qui auraient été observés si les distributions par diplôme avaient été égales, chez les deux sexes, à celles des hommes ou à celles des femmes (ce qui élimine l'effet de l'évolution de la dissimilarité) ont décliné au même rythme (Figure 1.3, courbes grises). L'évolution de la dissimilarité des distributions selon le sexe n'a donc pas joué de rôle important : de fait, on peut vérifier sur la figure 1.2 qu'elle a toujours été relativement faible, et qu'elle n'a guère évolué, puisque la massification scolaire a concerné les deux sexes de manière assez comparable.

Pour finir, notons que le taux d'endogamie attendu en situation de mise en couple aléatoire s'est stabilisé au cours de la décennie 2000, suite notamment à la disparition presque complète du groupe des titulaires du certificat d'études primaires et la stabilisation de la proportion de non diplômés, qui constituaient les deux sources principales de réduction de la concentration de la structure par diplôme de la population.

Figure 1.3 : Évolution du taux d'endogamie de diplôme des 30-59 ans attendu en situation de mise en couple aléatoire

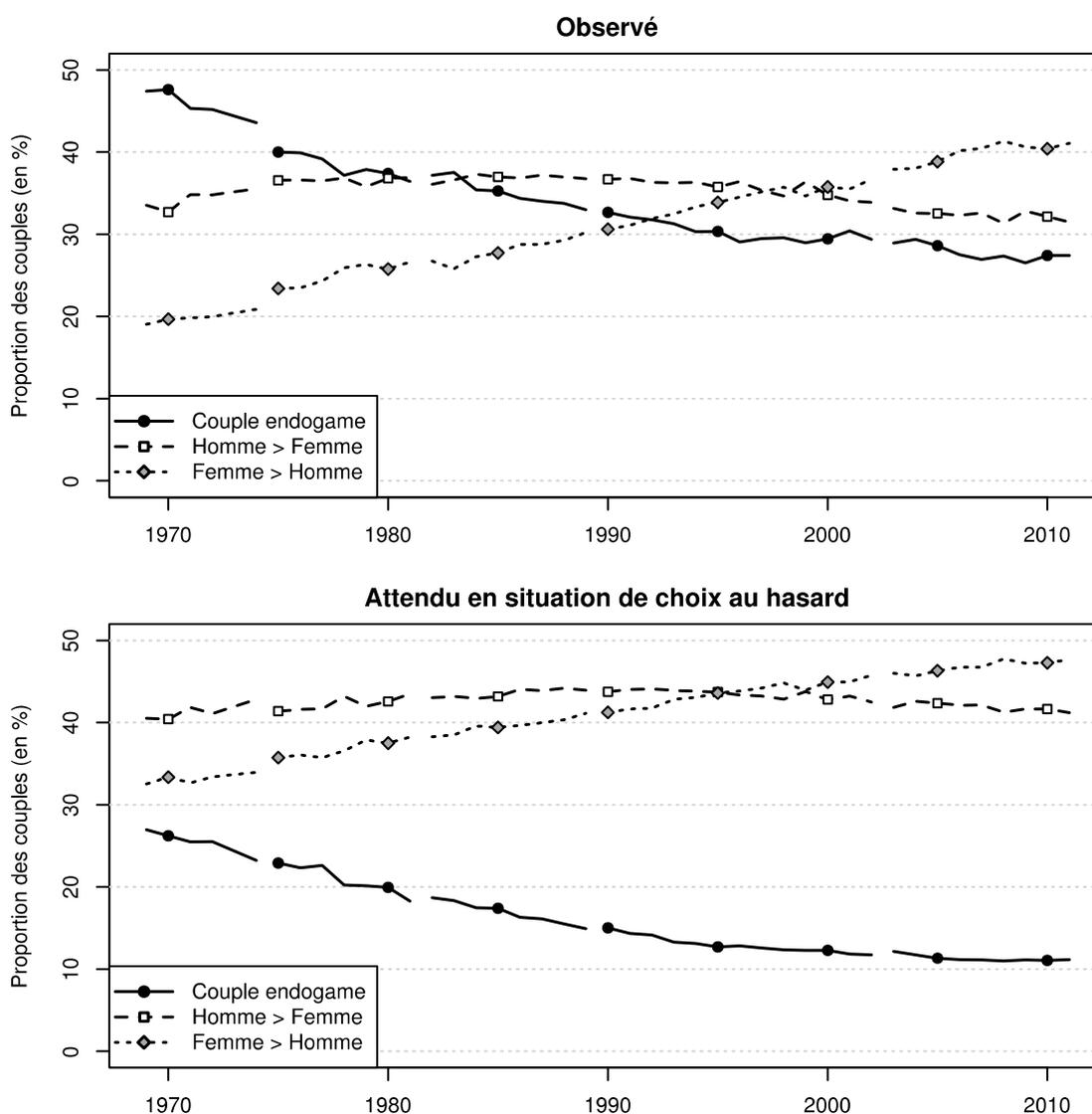


Source : enquêtes Emploi 1969-2011 (Insee)

Champ : individus âgés de 30 à 59 ans (célibataires inclus)

Les séries « Hommes » et « Femmes » représentent les taux qui seraient observés si les distributions étaient chez les deux sexes égales à celle observée pour le sexe en question, et mesurent la concentration de ces distributions. La série « Hommes et femmes » représente le taux attendu à partir de l'appariement aléatoire des hommes et des femmes. La différence entre ces deux types de courbes est liée à la dissimilarité des distributions observées chez les deux sexes.

Figure 1.4 : Taux d'endogamie, d'hypergamie et d'hypogamie de diplôme observés et attendus en situation de choix aléatoire du conjoint



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 30 à 59 ans ; les taux attendus sont calculés sur la base de la population totale des 30-59 ans, célibataires inclus

Note : le taux d'endogamie attendu en situation d'indépendance est le même que celui présenté dans la figure 1.3.

Cette évolution des taux attendus en situation d'indépendance s'est accompagnée d'une baisse parallèle du taux de couples endogames réellement observé, de 47 % en 1969 à 27 % en 2011 (Figure 1.4, graphique du haut). Ainsi, si nous observons une nette diminution de

l'endogamie de diplôme, nous ne pouvons affirmer à ce stade qu'elle est allée au-delà de ce que l'évolution de la structure de la population exigeait : en effet, le taux d'endogamie observé reste sur toute la période nettement supérieur à ce qui serait attendu, signe que l'endogamie relative persiste. Ce constat ne signifie pas que l'évolution observée n'a pas de conséquences tangibles et qu'elle peut être négligée : elle a au contraire profondément transformé la structure des relations entre groupes, notamment par un accroissement de la fréquence des configurations hétérogames. Nous y reviendrons plus bas.

Le second changement majeur tient à ce que les couples dans lesquels la femme est plus diplômée que son conjoint (hypogamie féminine) ont connu un développement régulier et rapide. À tel point qu'ils sont depuis l'an 2000 la configuration la plus fréquente, doublant leur part, de 20 % à 40 % (Guichard-Claudic, Testenoire et Trancart, 2009). Cette inversion des positions genrées au sein des couples est, au moins en partie, le résultat mécanique de l'élévation du niveau d'éducation des femmes décrite plus haut, ainsi que l'indique l'évolution parallèle du taux qui serait attendu en situation de choix aléatoire du conjoint (Figure 1.4, graphique du bas). La part de couples dans lesquels l'homme est plus diplômé que sa conjointe est, elle, restée plutôt stable, avec une légère augmentation jusqu'aux alentours de 1980, et une diminution un peu plus nette depuis. Cette évolution correspond elle aussi à ce qui serait attendu en situation d'indépendance.

L'asymétrie entre hypergamie et hypogamie féminines, dont les évolutions ne sont pas le reflet l'une de l'autre, s'explique par un effet plancher : le groupe des non diplômés était le groupe modal au début de la période, et on a vu qu'il a gardé la même taille chez les deux sexes. Ce sont donc principalement les chances des femmes de choisir un conjoint moins diplômé (option ouverte aux très diplômées) qui ont augmenté, plutôt que celles de choisir un conjoint plus diplômé (option ouverte aux peu diplômées) qui ont diminué. Nous reviendrons plus en détail sur la question de l'hypergamie au chapitre 4.

4.1.3 Un effondrement de la proportion de couples peu diplômés

Cette description de l'évolution de l'homogamie absolue resterait incomplète sans une analyse plus détaillée de la répartition des couples selon le diplôme des deux conjoints en début et en fin de période. Nous recourons pour cela à des représentations en mosaïque permettant de visualiser simultanément la structure des couples (homogamie absolue) et la

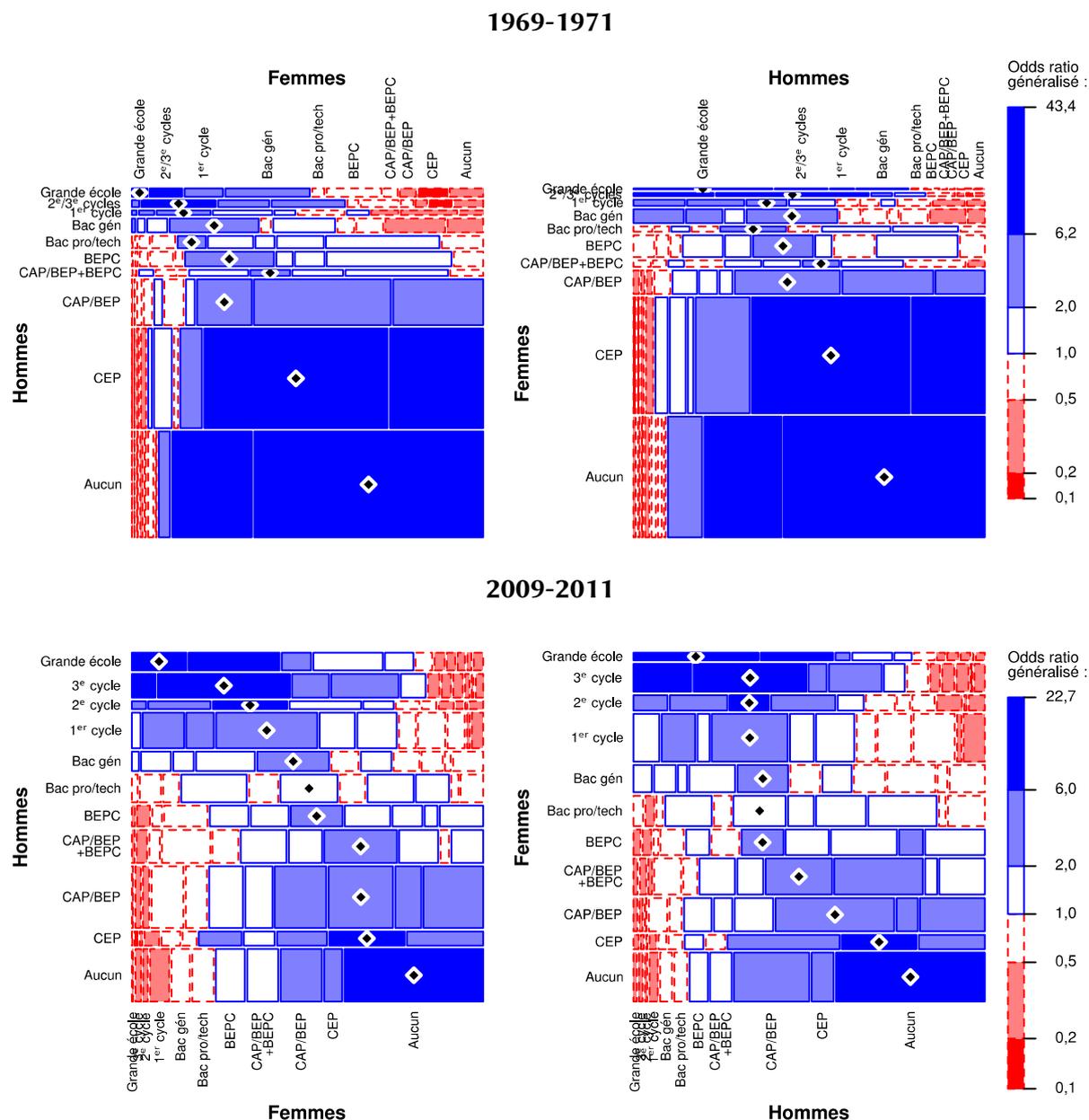
sur- ou sous-représentation des différentes configurations (homogamie relative). L'interprétation de ces graphiques est détaillée dans l'encadré 1.2. Cette représentation en mosaïque (Figure 1.5) fait immédiatement apparaître la profonde évolution de la structure des couples en termes de diplôme, au point que l'on peine à trouver des points communs entre les deux périodes.

En effet, autour de 1970, les quatre configurations dans lesquelles l'un ou l'autre des conjoints était soit non diplômé, soit titulaire du certificat d'études primaires (CEP) constituaient 60 % de l'ensemble des couples – ce qui était en grande partie le reflet de leur importance numérique dans la population. Ainsi que l'indiquent les *odds ratios* généralisés supérieurs à 6, ces configurations formaient en outre un bloc fortement homogame, nettement sur-représenté par rapport à ce que l'on attendrait d'une mise en couple au hasard. Un second pôle de forte homogamie était constitué par les diplômés du supérieur, qui représentait cependant une part très réduite de la population. Mais, malgré cette taille limitée, entre 20 et 40 % (selon les groupes) des hommes diplômés du supérieur prenaient une conjointe dans ce groupe, ce qui était même le cas de 45 à 75 % des femmes de ce niveau de diplôme.

Au bloc formé par les moins diplômés s'adjoignait le groupe des titulaires d'un CAP ou d'un BEP sans le BEPC, mais avec une association atténuée. Au-delà, entre les deux blocs des moins et des plus diplômés, s'établissait du BEPC (avec ou sans CAP ou BEP) au Baccalauréat une zone de moindre détermination du choix du conjoint. Si l'endogamie de ces groupes demeurait relativement forte (mais toujours plus faible qu'aux deux extrêmes), les situations d'exogamie se répartissaient sur un spectre assez large, couvrant une grande partie de l'échelle des diplômes avec des *odds ratios* inférieurs à 2.

Comme nous l'avons noté, les différences de structure par diplôme entre hommes et femmes étaient déjà relativement peu marquées, ce qui se retrouve dans la composition des couples. Si les femmes étaient encore globalement moins diplômées que les hommes, cette différence n'est pas flagrante ici : seule leur moindre possession du CAP et du BEP est véritablement apparente, les autres catégories étant de toute façon peu nombreuses même chez les hommes.

Figure 1.5 : Répartition des couples selon les diplômes des conjoints en 1969-71 et en 2009-11 (% en ligne)



Source : enquêtes Emploi (Insee)

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Lecture : En 1969-1971, 70 % des hommes non diplômés étaient en couple avec des femmes elles-mêmes non diplômées (graphique de gauche) ; inversement, 60 % des femmes non diplômées étaient en couple avec un homme non diplômé (graphique de droite). De même, le taux d'endogamie des hommes diplômés d'une grande école était de 5 %, alors que celui des femmes était de 40 %.

Cf. Encadré 1.2 (p. 100) pour des détails concernant la représentation en mosaïque.

Les données ne permettent pas de distinguer le deuxième et le troisième cycles universitaires avant 1982, ce qui n'est pas ici un problème étant donné la faible taille de ce groupe en 1969-1971. Du fait des faibles effectifs du premier cycle du général du supérieur, cette catégorie a ici été fusionnée avec le premier cycle professionnel pour améliorer la lisibilité.

Encadré 1.2 : La représentation des tables d'homogamie en mosaïque

Les graphiques en mosaïque (Hartigan et Kleiner, 1981 ; Friendly, 1994) consistent à représenter la structure d'une table de contingence sous la forme d'un rectangle divisé en cellules dont les aires sont proportionnelles aux effectifs. Appliquée à une table d'homogamie, cette méthode permet de saisir visuellement la répartition globale des couples entre les différentes configurations – opération malaisée à partir d'un tableau de pourcentages.

L'organisation des mosaïques retenue ici consiste à attribuer une hauteur commune à toutes les cellules d'une même ligne. Cette hauteur est proportionnelle à la part des individus du sexe placé en lignes appartenant à la catégorie considérée dans la population considérée. Au sein de chaque ligne, la largeur des cellules est alors proportionnelle à la part de chaque configuration au sein de la catégorie ligne : cette lecture correspond ainsi aux pourcentages en ligne. Pour chaque période, la figure de gauche prend pour référence (en lignes) les hommes, et celle de droite les femmes, autorisant des comparaisons directes entre les deux sexes. Les aires des cellules sont identiques dans les deux cellules, seule changeant la forme des rectangles. Afin de faciliter le repérage des cellules, un losange marque les cellules de la diagonale (configurations endogames), dont l'aire est généralement la plus importante de la ligne.

En complément de l'information transmise par les aires, qui correspond à l'homogamie absolue, les niveaux de couleurs des cellules reflètent les *odds ratios* généralisés de chaque configuration de couple. La combinaison de ces deux informations ne vise pas ici à permettre une analyse détaillée, mais à faire ressortir des ensembles de configurations associées (fond bleu avec un trait continu) ou opposées (fond rouge avec un trait pointillé), ainsi que des zones peu marquées du point de vue du choix du conjoint (fond blanc, correspondant à un *odds ratio* généralisé entre 0,5 et 2).

Notons que cette représentation diffère du chiasmogramme (Bozon et Héran, 1987b, 1988), dans lequel les aires correspondent aux proportions attendues en situation d'indépendance. La représentation en mosaïques met en avant la structure réellement observée de l'homogamie, au prix d'une figure plus complexe et non réversible entre lignes et colonnes. Une autre différence tient au recours à l'*odds ratio* généralisé, indépendant des marges, plutôt qu'à l'indice d'homogamie (cf. ci-dessus, p. 88), pour la coloration des cellules.

Ces graphiques en mosaïques ont été réalisés dans l'environnement R (R Core Team, 2014) à l'aide du paquet *vcd* (Meyer, Zeileis et Hornik, 2014)²⁵.

La composition des couples autour de l'année 2010 est bien différente en termes de proportions, même si l'association entre les groupes semble avoir conservé la même structure (point sur lequel nous reviendrons plus loin). Les effets de la diminution de la concentration de la structure par diplôme trouvent ici leur illustration concrète. Au bas de la hiérarchie des diplômes, le bloc constitué des non diplômés et des titulaires du certificat d'études n'a pas disparu, mais a vu sa taille divisée par six (10 % au lieu de 60 % en 1970). S'y associent toujours les titulaires d'un CAP ou d'un BEP, dont l'importance relative au sein de l'ensemble a mécaniquement crû (20 % des couples contre 70 % en 1970). Le bloc opposé, constitué des diplômés des deuxième et troisième cycles de l'université et des diplômés des grandes écoles, représente désormais 10 % des couples (contre 3 % en 1970), et même 20 % (contre 5 % en 1970) en lui associant l'ensemble des diplômés du supérieur avec qui l'association est cependant moins nette. Enfin, l'espace intermédiaire allant du BEPC au Baccalauréat est toujours visible, mais son importance a nettement augmenté : la proportion des couples dont au moins l'un des membres appartient à ces groupes est passée de 15 à 45 %.

Les cas de forte hétérogamie, pour lesquels la proportion observée est toujours bien inférieure à celle que l'on attendrait si les conjoints se mettaient en couple au hasard, se sont eux aussi fortement développés. Pour prendre l'exemple le plus frappant, la proportion des couples dans lesquels l'un des conjoints est diplômé du supérieur et l'autre possède un

25 Nous tenons à remercier Michael Friendly et David Meyer pour leur aide dans la mise au point de ces figures.

diplôme inférieur au Baccalauréat ou aucun diplôme a explosé, de 1 % à 8 % de l'ensemble des couples, passant donc d'une situation exceptionnelle à un cas relativement courant – quoique toujours atypique. Le statut social des diplômés du supérieur ne peut pas être resté le même à la suite d'une telle évolution.

Dans l'ensemble, on voit que l'homogamie de diplôme a été largement bouleversée par la massification scolaire. Si sa structure bipolaire demeure, elle a changé de signification avec l'égalisation des tailles des deux ensembles opposés et le développement de l'espace intermédiaire. Il est clair que la nouvelle structure est bien plus favorable aux mises en couples hétérogames que la précédente, du seul fait de la grande dispersion des individus entre des catégories de tailles limitées. Nous étudierons cette question plus en détail au chapitre suivant.

4.2 La classe sociale : déclin de l'indépendance et tertiarisation

4.2.1 Un déclin des indépendants et des ouvriers non qualifiés et une montée des cols blancs

La structure de la population française a aussi connu des transformations importantes du point de vue de la classe sociale. L'évolution majeure tient ici à l'explosion du taux d'activité des femmes, passant parmi les 30-59 ans de 47 % en 1969 à 85 % en 2011²⁶. La structure de la population active s'est elle aussi profondément transformée (Figure 1.6), prolongeant en grande partie des tendances à l'œuvre depuis au moins le début du XX^e siècle (Marchand et Thélot, 1997 ; Meron et Maruani, 2012). Des évolutions très proches sont observées lorsque le champ est élargi aux inactifs, en retenant leur dernière profession – quoiqu'il soit difficile de distinguer précisément les différences véritables entre actifs et inactifs de celles qui proviennent des difficultés de recodage.

On assiste tout d'abord au déclin régulier de deux ensembles : les indépendants et les ouvriers non qualifiés. Cette évolution est la plus nette chez les agriculteurs (IVc), dont la proportion parmi les actifs âgés de 30 à 59 ans passe entre 1969 et 2011 de 10 à 3 % chez les hommes, et même de 15 à 1 % chez les femmes. Chez ces dernières uniquement, la

26 Rappelons cependant que, si cette évolution marque une transformation majeure, « *les femmes ont toujours travaillé* », et les Françaises tout particulièrement (Schweitzer, 2002), les années 1950 et 1960 représentant de ce point de vue une parenthèse et non le signe d'une tradition intemporelle d'inactivité féminine (Meron et Maruani, 2012, chap. 2).

proportion de petits indépendants hors agriculture (IVb) a elle aussi nettement décru, passant de 13 à 3 %, l'augmentation de l'activité féminine s'étant réalisée par le salariat ; au contraire, moins élevée dès 1969, cette proportion est restée stable autour de 5 % chez les hommes. Finalement, le groupe des indépendants avec salariés (IVa) n'a pas vu sa taille évoluer très nettement : il représente toujours environ 5 % des hommes et 2 % des femmes.

Du côté des salariés, ce sont les professions les moins qualifiées qui ont connu un déclin. Les ouvriers non qualifiés hors agriculture (VIIa) constituaient, à plus de 20 %, le premier groupe parmi les hommes en 1969 ; il n'est plus que le troisième en 2011, avec 15 % des effectifs. La tendance est encore plus claire chez les femmes, chez qui l'on passe de 14 à 6 %. Cette évolution représente la seule véritable rupture de tendance au cours du XX^e siècle. Quant aux ouvriers agricoles (VIIb), qui représentaient déjà seulement 3 % des hommes et 1 % des femmes en 1969, ils ont presque complètement disparu.

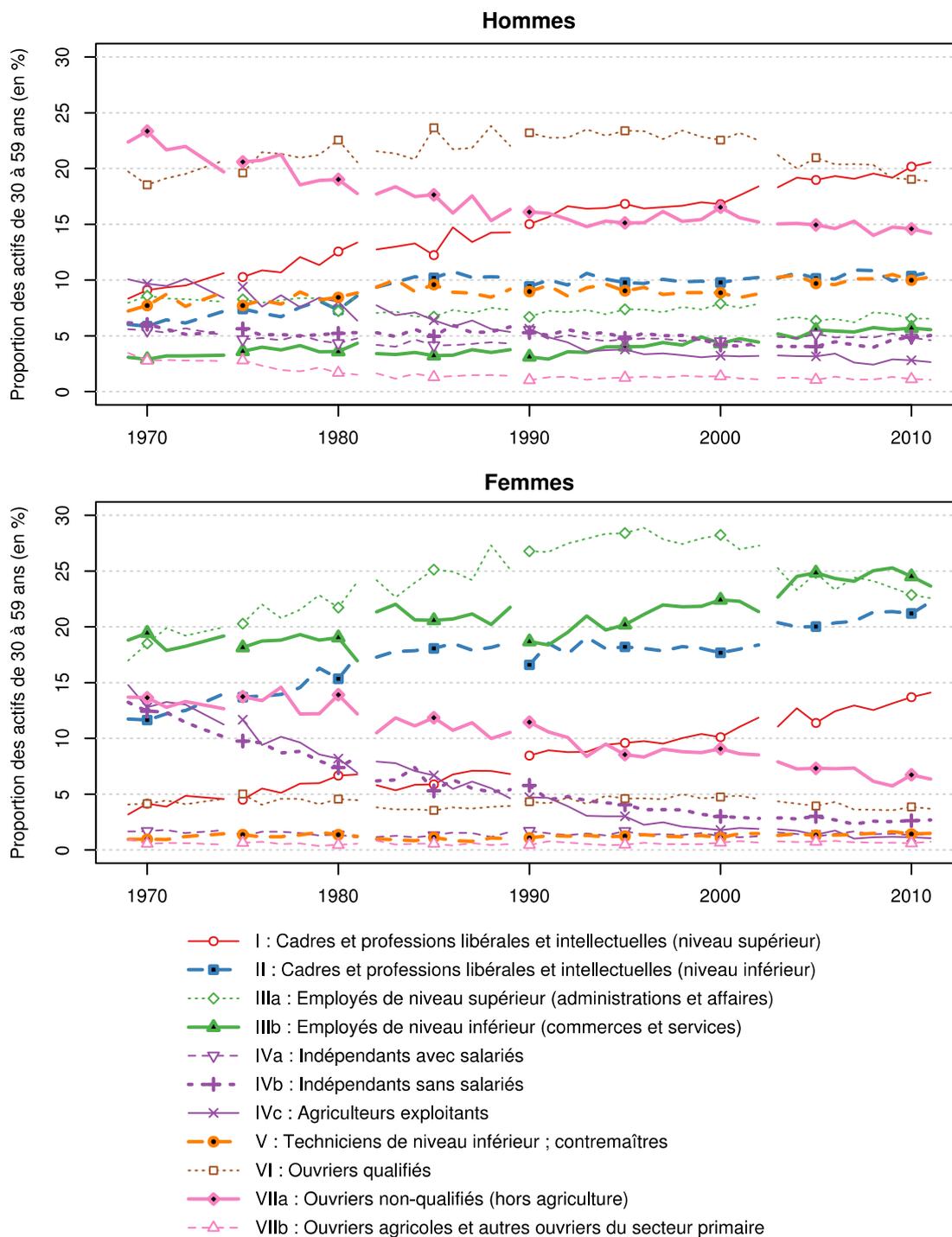
Le déclin de ces groupes s'est fait au profit des catégories salariées en col blanc. Au sommet de la hiérarchie, la croissance numérique des catégories intermédiaires et supérieures (*service class*, I et II) est frappante : ce groupe passe de 15 % de la population active chez les deux sexes en 1969 à 30 % chez les hommes et même 36 % chez les femmes en 2011. Dans l'ensemble, la structure sociale a ainsi connu un mouvement ascendant très net. Néanmoins, la part des emplois les plus qualifiés (I) reste nettement plus élevée chez les hommes (20 %) que chez les femmes (15 %).

Les différences entre hommes et femmes sont encore plus visibles parmi les classes populaires. Les postes d'employées qualifiées (IIIa) et non qualifiées (IIIb) se sont développés rapidement chez les femmes, passant de 35 à 46 % de la population active²⁷ ; en revanche, leur part est restée plutôt stable chez les hommes, autour de 11 %. De leur côté, les ouvriers représentaient 45 % des hommes et 20 % des femmes en 1969, contre 35 % et 10 % en 2011. Contrepartie de la croissance des postes d'employées, le déclin des emplois ouvriers a été plus fort chez les femmes, qui occupaient plus souvent des emplois non qualifiés. La part des ouvriers qualifiés (VI), qui est restée très stable au cours de la période, est nettement plus forte chez les hommes (de 20 à 25 %) que chez les femmes (4 à 5 %). De même, la proportion

27 On peut noter une certaine instabilité des frontières entre employées non qualifiées (IIIb), qualifiées (IIIa) et cadres et professions intellectuelles de rang inférieur (II) selon les séries. Celle-ci se jouant au niveau des professions détaillées, elle n'a pas pu être corrigée (*cf.* la section sur le recodage de la classe EGP, p. 72).

de techniciens de niveau inférieur et de contremaîtres (V) est passée de 7 à 10 % chez les hommes, alors qu'elle est restée inférieure à 2 % chez les femmes. Au total, suite à ces évolutions divergentes, la ségrégation genrée des professions s'est nettement renforcée au sein des classes populaires – du moins au niveau agrégé qui est ici le nôtre (Meron, Okba et Viney, 2006).

Figure 1.6 : Évolution de la structure par classe sociale des actifs âgés de 30 à 59 ans



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : actifs âgés de 30 à 59 ans

4.2.2 Une nette diminution du taux d'endogamie de classe sociale

La comparaison des structures par classe sociale des hommes et des femmes révèle donc un tableau bien différent de celui déjà décrit concernant le diplôme. En effet, la concentration des individus entre les différentes classes sociales est bien plus faible au début de la période que celle suivant le diplôme ; surtout, elle est restée stable chez les hommes, et a sensiblement augmenté chez les femmes (Figure 1.7, courbes grises). Chez ces dernières, cette évolution s'explique par le développement des catégories déjà fortement représentées²⁸ : les employées (IIIab) et les cadres et professions intellectuelles de rang inférieur (II). Chez les hommes, la stabilité est le résultat de mouvements importants de directions opposées, dont notamment le remplacement, en position de catégorie la plus représentée, du groupe des ouvriers non qualifiés (VIIa) par celui des cadres et professions intellectuelles supérieures (I).

Mais si la concentration de la distribution a nettement augmenté chez les femmes, les distributions des deux sexes se sont aussi par le même mouvement éloignées l'une de l'autre (un mouvement déjà relevé par Vallet, 1986, p. 671). Au total, le taux d'endogamie attendu en situation de mise en couple aléatoire n'a donc que très légèrement augmenté, passant de 8 à 9,5 % (Figure 1.7, courbe noire). Le contraste avec la situation relevée à propos de l'homogamie de diplôme est frappant, si ce n'est que les taux correspondant à ces deux composantes de l'homogamie se sont rapprochés dans la période récente. Doit-on pour autant s'attendre à une augmentation du taux d'endogamie de classe sociale réellement observé ?

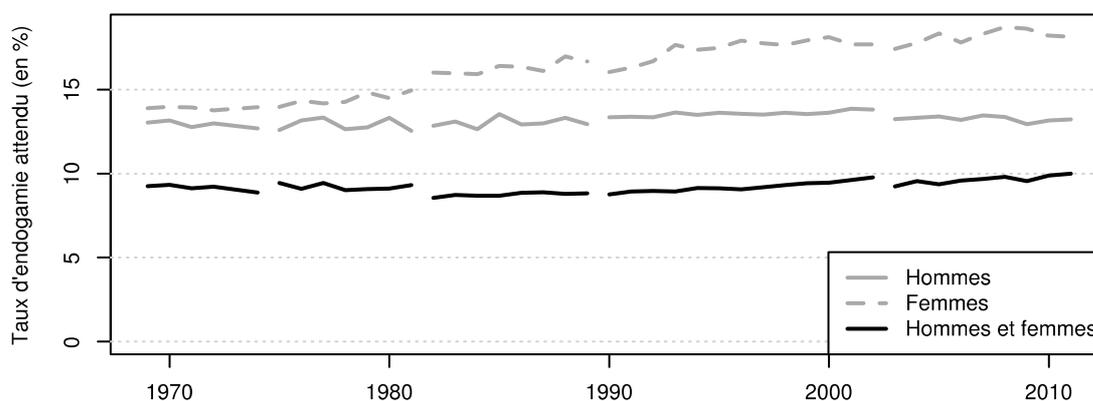
Puisque l'endogamie attendue n'a pas évolué au cours du temps, toute évolution de l'endogamie réellement observée correspond à une évolution du choix du conjoint par-delà l'effet des contraintes structurelles²⁹. Or, on observe une nette baisse de la part de couples endogames : ceux-ci passent de 30 % à 20 % de l'ensemble des couples entre 1969 et 2011, et même de 40 % à 22 % des couples bi-actifs. Cette tendance est encore plus marquée si l'on considère employés et ouvriers (IIIab, VI, VIIab) comme formant un seul et même groupe :

28 Cette plus forte concentration de la structure par classe sociale chez les femmes peut aussi s'interpréter comme la conséquence de la moins bonne représentation des positions féminines par les classifications des professions (Amossé, 2004).

29 Nous ne présentons pas ici les taux d'hypergamie et d'hypogamie, dont le calcul exige de se doter d'un ordre hiérarchique des classes sociales, qui n'est pas évident dans tous les cas. Cette question sera traitée au chapitre 4.

dans ce cas, la proportion de couples endogames tombe de 63 % à 47 % de l'ensemble des couples. Contrairement au cas du diplôme, nous sommes donc en mesure d'affirmer dès maintenant que l'endogamie relative de classe sociale a diminué, dans des proportions qui restent cependant à déterminer plus précisément³⁰.

Figure 1.7 : Évolution du taux d'endogamie de classe sociale des actifs de 30 à 59 ans attendu en situation de mise en couple aléatoire



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : actifs âgés de 30 à 59 ans (célibataires inclus)

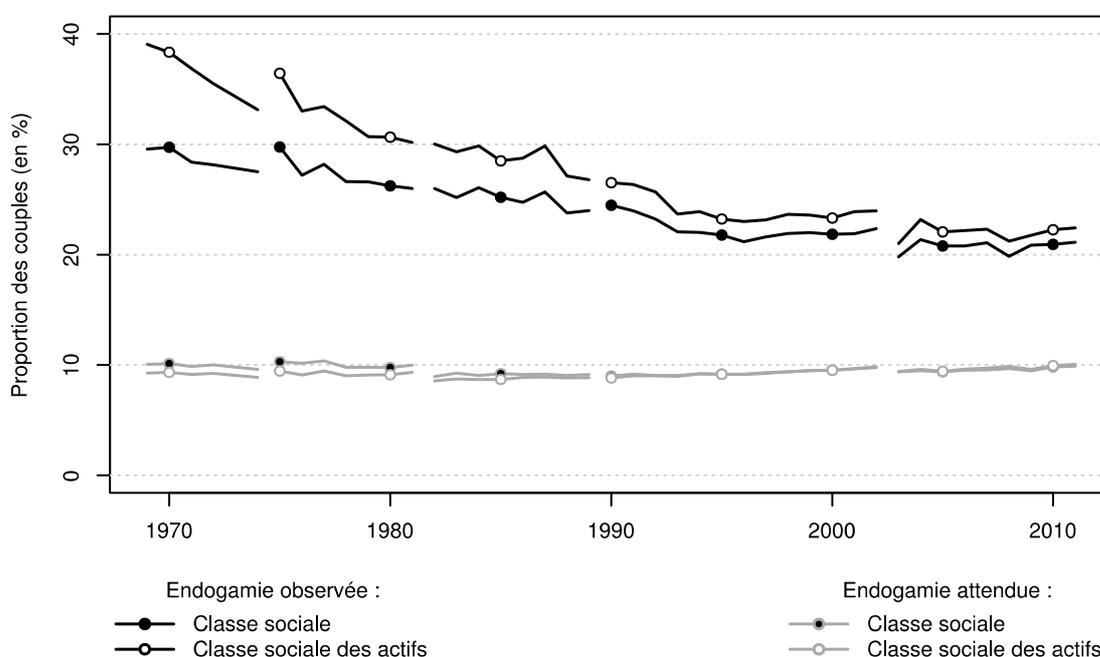
Les séries « Hommes » et « Femmes » représentent les taux qui seraient observés si les distributions étaient chez les deux sexes égales à celle observée pour le sexe en question, et mesurent la concentration de ces distributions. La série « Hommes et femmes » représente le taux attendu à partir de l'appariement aléatoire des hommes et des femmes. La différence entre ces deux types de courbes est liée à la dissimilarité des distributions observées chez les deux sexes.

Ce déclin est visible chez les actifs aussi bien que chez les inactifs, ce qui exclut qu'il soit dû à l'entrée massive des femmes sur le marché du travail sur la période. Néanmoins, on peut remarquer que les couples de bi-actifs étaient nettement plus endogames que les autres au début de la période – différence qui s'est mécaniquement réduite avec le développement de l'activité féminine, les premiers représentant désormais la grande majorité de l'ensemble. La plus forte homogamie des couples bi-actifs peut s'interpréter comme la confirmation du fait que la profession est un meilleur indicateur de la position sociale des femmes actives

³⁰ En effet, de même que l'indice d'homogamie critiqué plus haut avec lequel il partage une certaine similarité, ni le rapport ni l'écart entre le taux d'endogamie observé et le taux attendu ne sont des indicateurs indépendants des marges de la table.

que des inactives. Mais aussi potentiellement du fait que le codage de la profession des actifs est plus précis que celui de la dernière profession des inactifs ; ce phénomène explique vraisemblablement une partie assez faible de l'écart, comme on peut le voir au fait que la qualité bien moindre de la seconde variable dans la série 1975-1981 (*cf.* p. 76) n'a aucun effet visible.

Figure 1.8 : Taux d'endogamie de classe sociale observés et attendus en situation de choix aléatoire du conjoint



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 30 à 59 ans et les deux conjoints ont déjà travaillé ; les taux attendus sont calculés sur la base de la population totale des 30-59 ans ayant déjà travaillé, célibataires inclus

Note : le taux d'endogamie des actifs attendu en situation d'indépendance est le même que celui présenté dans la figure 1.7.

4.2.3 Un déclin des couples d'indépendants et une perte d'unité des classes populaires

Malgré l'ampleur de la baisse du taux d'endogamie de classe sociale, les transformations de la structure des couples sont restées plus limitées que dans le cas du diplôme (Figure 1.9). L'évolution la plus évidente tient au déclin du bloc constitué par les indépendants. Toujours

fortement endogames, ces groupes ont cependant nettement perdu en importance : la part des couples (actifs et inactifs confondus) associant deux agriculteurs (IVc) s'est effondrée de 15 à 1 %, et celle des couples associant deux indépendants (IVab) est tombée de 6 à 1 %. La diminution de la part de ces deux groupes dans la population active a entraîné une nette diminution de leur taux d'endogamie, particulièrement claire chez les hommes agriculteurs, où il passe de 87 à 32 %³¹.

À l'inverse, on peut noter le fort développement d'un pôle constitué par les classes intermédiaires et supérieures (I, II). Ce développement est le résultat de l'augmentation des effectifs de ces catégories chez les deux sexes, mais il ne s'est pas accompagné d'une plus forte endogamie relative de ces groupes – constat que l'on peut attribuer à la croissance de l'hétérogénéité interne de l'ensemble constitué par les cadres (Bouffartigue, 2001). En revanche, il est intéressant de relever que l'association entre les hommes de ce groupe et les employées qualifiées (IIIa) a légèrement diminué en termes d'*odds ratios* : il semble que l'apparition d'une sur-représentation des femmes parmi les classes supérieures a entamé le schéma classique de l'hypergamie féminine selon la profession – point sur lequel nous reviendrons au chapitre 4.

Du fait du net déclin des emplois ouvriers féminins, les couples endogames ouvrier-ouvrière (VI, VIIab) ont nettement perdu en importance, passant de 18 % de l'ensemble des couples au début de la période à seulement 7 %. Si 60 % des ouvrières sont en couple avec un ouvrier en 2009-2011, soit à peine moins que les 68 % constatés en 1969-1971, c'est le cas de seulement 21 % des ouvriers – contre tout de même 43 % au début de la période. L'augmentation de la ségrégation genrée au sein des classes populaires présente ainsi des effets bien visibles sur la composition des couples. Si l'on retient une définition plus large des classes populaires, on observe que la proportion d'hommes ouvriers ou employés (qualifiés ou non, soit IIIab, VI, VIIab) qui sont en couple avec une femme du même groupe a augmenté de 79 à 85 %, alors qu'elle a au contraire diminué chez les femmes, passant de 71 à 59 %. Cette configuration n'a que légèrement perdu en importance, passant de 46 à 39 % du total des couples.

31 Ces taux cachent des différences importantes à l'intérieur du groupe : l'endogamie des petits agriculteurs, peu spécialisés et peu diplômés, est plus forte que celle des agriculteurs fortement spécialisés et diplômés, qui, employant des salariés, associent moins fréquemment leur conjointe à leur exploitation (Giraud et Rémy, 2008).

La tendance opposée entre hommes et femmes s'explique par le fait que ces dernières sont moins souvent ouvrières non qualifiées (VIIab) qu'auparavant, et plus souvent employées (IIIab) : alors que le premier groupe est assez fortement associé aux catégories populaires, le second prend ses conjoints de manière bien moins contrastée sur l'ensemble de l'espace social³². On voit ainsi que la tertiarisation des emplois féminins ne représente pas qu'un changement de dénomination sans conséquences sur la structure sociale : elle est à l'origine de l'affaiblissement de la cohésion des classes populaires.

En outre, le groupe des employés ne forme pas non plus un ensemble homogène du point de vue du choix du conjoint (Bozon, 1991b). À toutes les périodes, la différence entre employés qualifiés (IIIa) et non qualifiés (IIIb) est bien visible, les premiers prenant des conjoints aussi bien parmi les catégories intermédiaires et supérieures que parmi les catégories populaires, alors que les seconds sont plus nettement associés aux catégories populaires. Si, tirant la conséquence de ces choix conjugaux, on définit les classes populaires comme l'ensemble constitué par les seuls ouvriers (VI, VIIab) et les employés non qualifiés (IIIb), on observe un déclin plus net que précédemment des couples endogames à ce groupe : de 33 % en 1969-1971 à 22 % en 2009-2011, soit une diminution d'un tiers. La signification de l'évolution de la structure par classe sociale n'apparaît ainsi pleinement qu'en recourant à une classification assez fine des professions – et notamment des professions féminines.

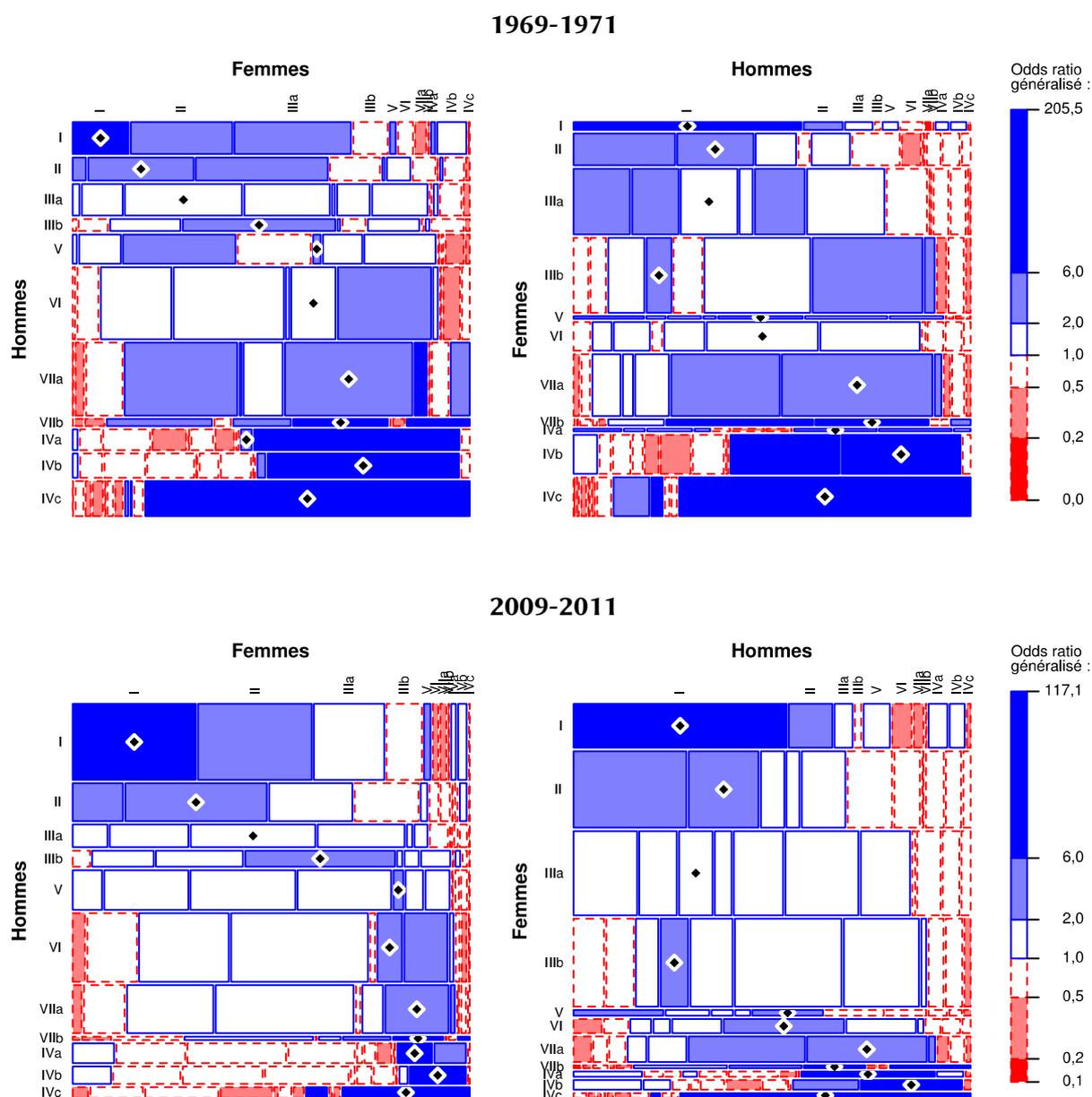
Ce mouvement n'affecte pas que les groupes d'employés. Si les ouvriers non qualifiés (VIIa) se distinguaient nettement des ouvriers qualifiés (VI) par une plus forte association aux classes populaires au début de la période, ces deux groupes présentent une association moins marquée chez les hommes dans les années récentes. Cette évolution fait écho à l'affaiblissement de la « *capacité de préservation de l'entre-soi* » de la classe ouvrière et à la remise en cause de la fameuse l'opposition « eux »/« nous » théorisée par Richard Hoggart, qui font suite à la désindustrialisation, à la tertiarisation des emplois, à la massification scolaire ou encore à l'augmentation de la précarité (Mauger, 2006, 2013, p. 253).

Comme nous l'avons déjà souligné plus haut, la baisse du taux d'endogamie semble donc s'être accompagnée chez les deux sexes d'un déclin de l'endogamie relative des différentes classes sociales, dont nous préciserons l'ampleur au chapitre suivant. Cette

32 Cette évolution ne semble pas un simple effet de nomenclature. En effet, comme chez les ouvrières, nous distinguons ici employées qualifiées et non qualifiées : la tertiarisation ne se traduit donc pas par un codage moins détaillé des professions des femmes.

évolution se lit aussi au développement des configurations nettement hétérogames : la proportion de couples associant un membre des cadres et professions intellectuelles supérieures (I) et un membre des classes populaires au sens large (IIIab, VI et VIIab) a presque triplé passant de 3 % en 1969 à 8 % de l'ensemble des couples en 2011. Si nous nous concentrons sur la situation encore plus atypique associant un membre des cadres et professions intellectuelles supérieures (I) et un ouvrier ou employé non qualifié (IIIb et VIIab), nous relevons un triplement de leur part, de 1 % à 3 % des couples. Ainsi, bien que la composition des couples en termes de classes sociales n'ait pas été radicalement bouleversée comme du point de vue du diplôme, des évolutions importantes ont cependant pris place.

Figure 1.9 : Répartition des couples selon la classe EGP des conjoints en 1969-71 et en 2009-11 (% en ligne)



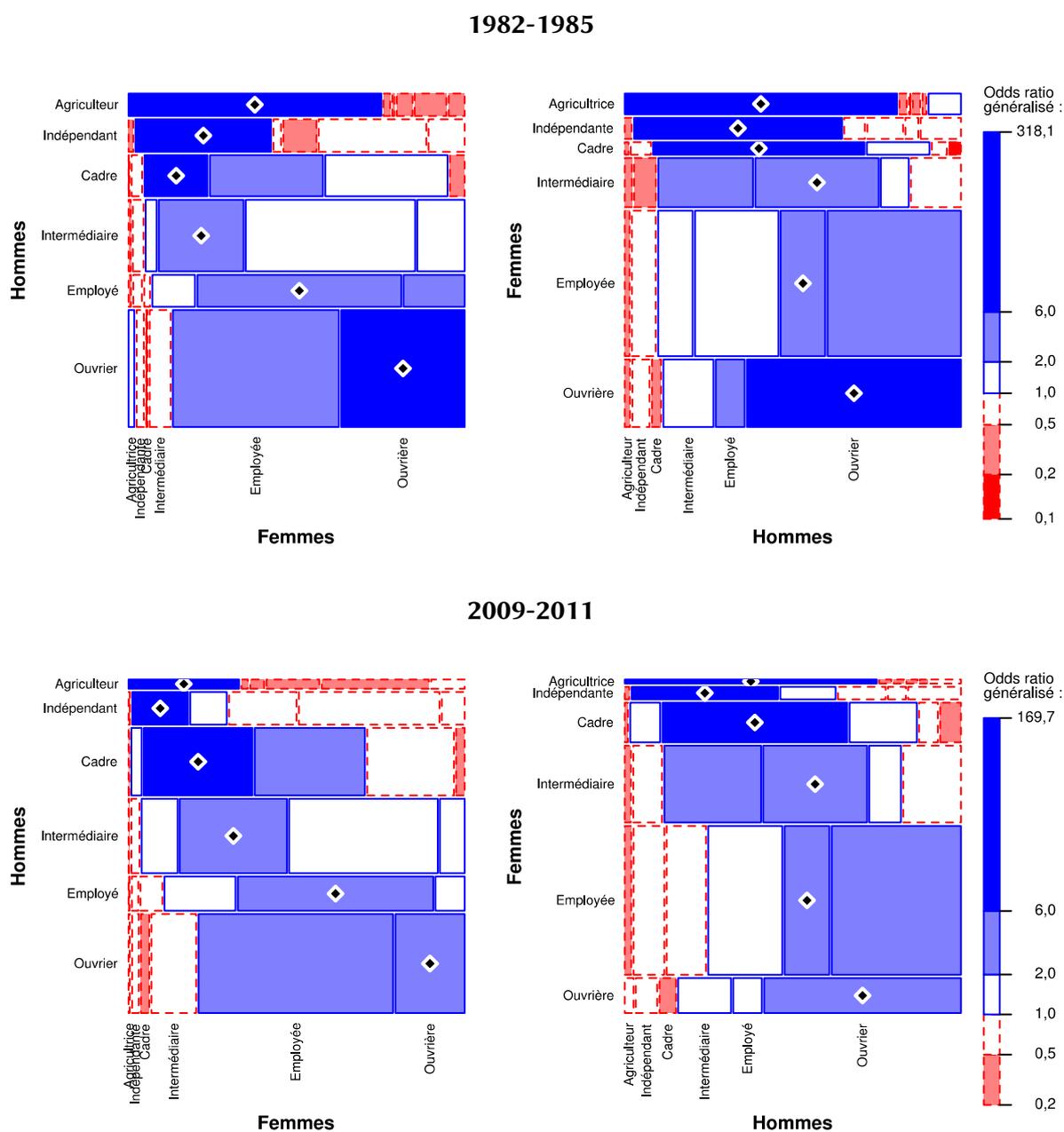
Source : enquêtes Emploi (Insee)

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans et les deux conjoints ont déjà travaillé (actifs et inactifs confondus)

Cf. Encadré 1.2 (p. 100) pour des détails concernant la représentation en mosaïque, et Figure 1.6 (p. 105) pour l'intitulé des catégories.

Lecture : En 1969-1971, 15 % des hommes cadres et professions intellectuelles supérieures (I) étaient en couple avec des femmes du même groupe (graphique de gauche) ; c'était le cas de 61 % des femmes de ce groupe (graphique de droite).

Figure 1.10 : Répartition des couples selon la catégorie socioprofessionnelle des conjoints en 1982-85 et en 2009-11 (% en ligne)



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans et les deux conjoints ont déjà travaillé

Cf. Encadré 1.2 (p. 100) pour des détails concernant la représentation en mosaïque.

Lecture : En 1982-1985, 51 % des ouvriers étaient en couple avec une employée (graphique de gauche) ; inversement, 41 % des employées étaient en couple avec un ouvrier (graphique de droite).

4.2.4 La structure de l'homogamie vue à travers la PCS

Afin de confirmer les observations réalisées à l'aide de la classe EGP des conjoints dans une nomenclature plus couramment utilisée dans le cas français, nous présentons dans la figure 1.10 l'homogamie en termes de catégorie socioprofessionnelle des conjoints. Moins détaillé, le premier niveau de la nomenclature des PCS permet aussi une lecture plus facile. Comme annoncé plus haut, nous n'avons pas procédé au recodage de la PCS avant 1982 ; aussi, la comparaison est-elle limitée aux trente dernières années, ce qui implique que les évolutions entre les deux tables sont moins marquées.

L'ensemble des tendances déjà évoquées se retrouvent. La principale différence tient aux professions intermédiaires, qui, à cheval sur le bas de la *service class* et le haut des employés de la classification EGP, apparaissent encore plus centrales, justifiant pleinement leur appellation du point de vue du choix du conjoint.

4.3 L'origine sociale : une image en différé

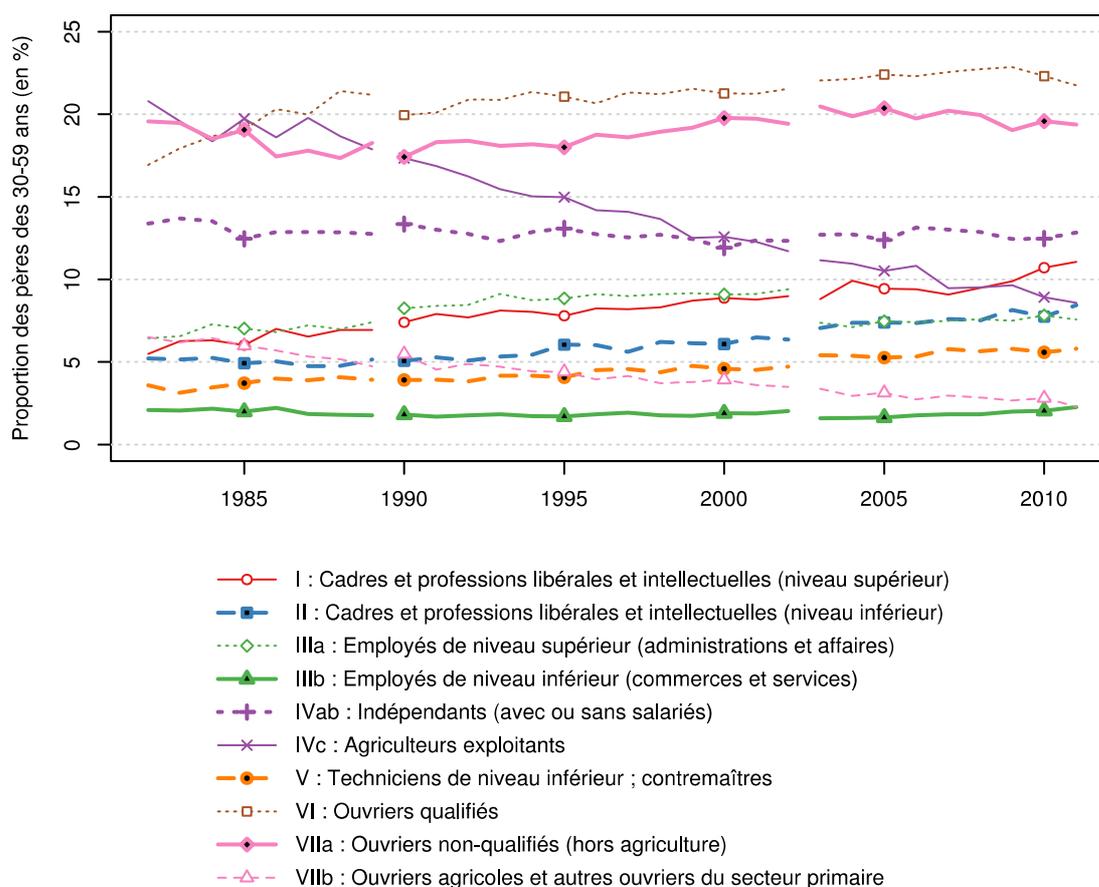
4.3.1 Des origines moins souvent agricoles, et plus souvent en col blanc

Pour finir, nous portons notre attention sur l'homogamie d'origine sociale, approchée ici par la classe EGP des pères des conjoints. Cette dernière composante de l'homogamie présente l'intérêt de ne pas faire jouer des différences du point de vue de la structure de la population des hommes et des femmes : par définition, les origines ne sont pas différenciées selon le sexe. De ce fait, on peut s'attendre à une plus grande symétrie dans le choix du conjoint que du point de vue des variables acquises comme le diplôme ou la classe sociale – même si certains phénomènes fortement genrés comme l'exode rural peuvent introduire des asymétries.

L'origine sociale des individus âgés de 30 à 59 ans (hommes et femmes confondus), qui n'est disponible dans les enquêtes que depuis 1982, a évolué plus lentement que la classe sociale des conjoints (Figure 1.11) – ce qui reflète principalement les évolutions antérieures de la population active, avec un retard d'une génération. Nous pouvons encore une fois observer un net déclin de la proportion d'agriculteurs (IVc), qui passe de 20 à 9 %, ainsi que d'ouvriers agricoles (VIIb), qui passe de 6 à 2 %. En revanche, la proportion d'ouvriers (VI et VIIa) augmente encore légèrement dans la génération des pères, et la proportion

d'indépendants (IVab) est stable autour de 13 %, alors que ces deux groupes voient leurs effectifs diminuer dans la génération des fils (Figure 1.6, p. 105). Les autres professions salariées se sont développées ; en particulier, la proportion des pères membres des classes intermédiaires et supérieures (I et II) augmente de 11 à 19 %³³.

Figure 1.11 : Évolution de la structure par classe sociale d'origine des 30-59 ans



Source : enquêtes Emploi 1982-2011 (Insee)

Champ : individus âgés de 30 à 59 ans

4.3.2 Une diminution du taux d'endogamie d'origine sociale

La concentration de la distribution des origines sociales est restée très stable depuis 1982 (Figure 1.12, courbe grise), et légèrement supérieure à celle de la distribution des fils

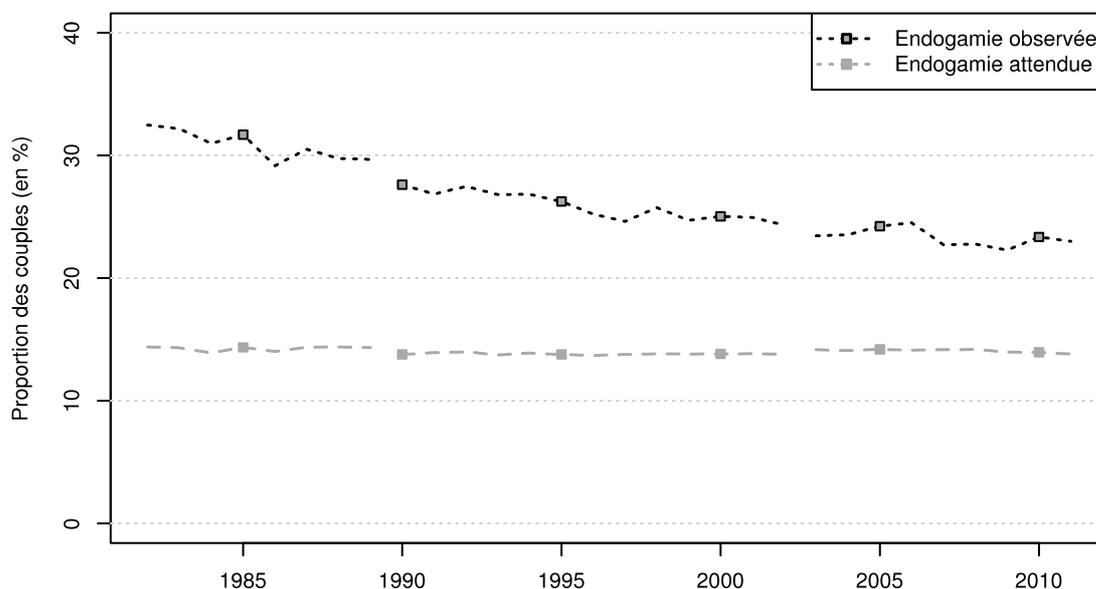
33 On retrouve en 2003 une rupture dans le codage des cadres et professions intellectuelles de rang inférieur (II) et des employés qualifiés (IIIa), que l'on avait déjà relevée concernant la profession des conjoints (note 7, p. 65).

observée dans la figure 1.7, p. 107). Cette stabilité est, ici encore, le résultat de mouvements divergents, dont le déclin de la part des agriculteurs (IVc, qui représentaient le groupe le plus important en 1982), compensé par une augmentation de la proportion d'ouvriers qualifiés (IVc) et de cadres et professions intellectuelles (I et II). Étant donné que les distributions des origines sociales sont identiques chez les hommes et les femmes (dissimilarité nulle)³⁴, aucun autre facteur ne vient affecter le taux d'endogamie attendu en situation de mise en couple aléatoire. Comme dans le cas de la classe sociale, toute évolution du taux d'endogamie observé indique donc une modification dans le même sens de l'endogamie relative.

Or, on note ici encore une claire diminution du taux d'endogamie d'origine sociale, de 33 % en 1982 à 23 % en 2011 (Figure 1.12, courbe noire). Ce mouvement, qui s'est produit sans que les contraintes structurelles y prennent aucune part, signale une modification importante des déterminants du choix du conjoint, dont nous détaillerons l'ampleur au chapitre suivant.

34 Cette identité, qui est évidemment imparfaite, est due à l'absence de l'association entre statut social des parents et sexe des enfants. Des différences entre hommes et femmes dans les taux de célibat selon l'origine sociale existent cependant ; cette dimension sera introduite au chapitre 4.

Figure 1.12 : Taux d'endogamie d'origine sociale observés et attendus en situation de choix aléatoire du conjoint



Source : enquêtes Emploi 1982-2011 (Insee)

Champ : couples dans lesquels l'un des conjoints est âgé de 30 à 59 ans ; les taux attendus sont calculés sur la base de la population complète des 30-59 ans, célibataires inclus

4.3.3 Une lente évolution structurelle et un affaiblissement de l'homogamie

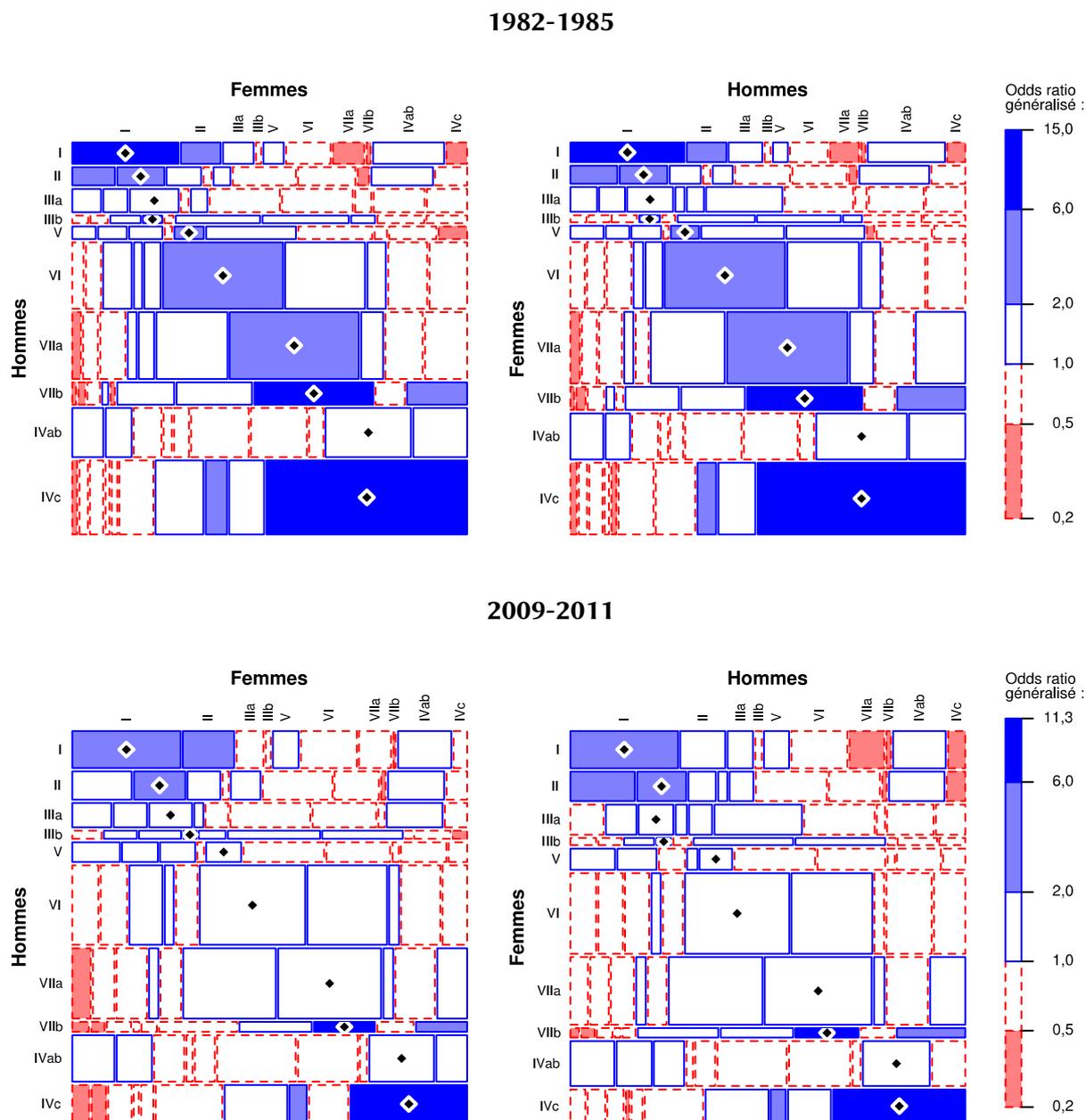
Les évolutions dans la structure de la population par origine sociale depuis 1982 ayant été plus limitées que celle par classe sociale et par diplôme depuis 1969, la transformation de la structure des couples est aussi de moindre ampleur suivant cette composante de l'homogamie (Figure 1.13). La différence principale tient ici à la réduction de la proportion des couples formés de deux enfants d'agriculteur (IVc), passant de 11 à seulement 3 % entre 1982-1985 et 2009-2011. Le pôle constitué par les classes intermédiaires et supérieures (I et II) gagne légèrement en importance, de 3 à 7 % de l'ensemble des couples, mais voit son endogamie relative diminuer légèrement.

Au centre, un bloc représentant un tiers des couples regroupe les catégories populaires au sens large (employés et ouvriers qualifiés ou non, IIIab, VI, VIIab, auxquels on peut rattacher les techniciens et contremaîtres, V). Si sa taille est extrêmement stable dans le temps, on observe clairement une diminution de l'endogamie relative de chacune des

catégories le composant. Surtout, ce bloc uni en 1982-1985 a perdu sa cohérence en 2009-2011 : l'association entre enfants de techniciens et de contremaîtres (V) et enfants d'ouvriers qualifiés (VI) est devenue légèrement négative, alors que les enfants d'ouvriers non qualifiés (VIIb) ne se trouvant même positivement associés qu'avec les enfants d'ouvriers non qualifiés et d'agriculteurs. Cette transformation, que l'on n'observe pas du point de vue de la classe sociale des conjoints, semble marquer un certain éclatement des classes populaires au fil des générations. On peut relier ce mouvement aux évolutions importantes déjà notées concernant la structure par diplôme et par classe sociale de la population : des individus d'origines sociales légèrement différentes au regard des destins sociaux qui étaient par le passé les plus probables se sont trouvés éloignés par une diversification des « chances de vie » liées à la massification scolaire et à l'expansion par le haut de la distribution des professions.

En conclusion, on peut noter que les différences entre hommes et femmes apparaissent ici assez peu marquées, ce qui indique que le phénomène de l'hypergamie féminine est relativement limité du point de vue de l'origine sociale. Or, il s'agit de la seule composante de l'homogamie étudiée qui ne présente pas de différences structurelles entre hommes et femmes : toute différence genrée serait l'indice d'une préférence des individus ou d'un effet des normes de genre concernant le choix du conjoint, plutôt que d'autre processus sociaux liés à l'éducation ou au marché du travail. Cette question sera approfondie au chapitre 4.

Figure 1.13 : Répartition des couples selon la classe EGP d'origine des conjoints en 1982-85 et en 2009-11 (% en ligne)



Source : enquêtes Emploi (Insee)

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Cf. Encadré 1.2 (p. 100) pour des détails concernant la représentation en mosaïque, et Figure 1.11 (p. 115) pour l'intitulé des catégories.

Lecture : En 1982-1985, 28 % des hommes dont le père était membre du groupe cadres et professions intellectuelles supérieures (I) étaient en couple avec des femmes de la même origine (graphique de gauche) ; c'était le cas de 31 % des femmes de cette origine (graphique de droite).

5 Conclusion

La première approche des évolutions de l'homogamie de diplôme, de classe et d'origine sociales développée dans ce chapitre a mis en évidence une baisse assez nette de la ressemblance entre conjoints selon les trois dimensions étudiées en France entre 1969 et 2011. Parmi les couples cohabitants, mariés ou non, dans lesquels l'un des conjoints est âgé de 30 à 59 ans, la proportion de couples endogames (ceux dans lesquels, respectivement, les deux conjoints possèdent le même diplôme, appartiennent à la même classe sociale ou sont issus de la même classe sociale) a baissé de 30 à 50 %, selon la composante retenue.

Dans le cas du diplôme, cette baisse est en partie un effet mécanique de la massification scolaire, qui a fortement réduit la probabilité qu'un homme et une femme pris au hasard aient exactement le même diplôme. Alors que plus de la moitié des couples associaient en 1969 deux individus ayant au plus le certificat d'études, ce n'est le cas que d'un couple sur dix en 2011. À l'inverse, les cas de très forte hétérogamie, s'ils restent rares, se sont nettement développés. En outre, du fait de la récente supériorité des femmes en termes de niveau d'éducation, les couples dans lesquels la femme est plus diplômée que son conjoint sont devenus plus fréquents aux alentours de l'an 2000.

L'endogamie de classe sociale a elle aussi nettement décliné. Mais, contrairement à celle de diplôme, son évolution a pris place en dehors de toute influence des évolutions structurelles : elle reflète donc clairement une moindre détermination sociale du choix du conjoint, sur laquelle nous reviendrons au chapitre suivant. Elle trouve en partie son explication dans la diminution de la proportion d'indépendants (et notamment d'agriculteurs), groupes fortement endogames, et le développement de groupes aux choix conjugaux moins marqués, comme les employés ou les cadres et professions intellectuelles de rang inférieur. L'ensemble formé par les ouvriers et employés a lui aussi perdu en cohésion. Enfin, ici encore, les configurations très hétérogames associant un membre des classes supérieures à un membre des classes populaires, qui restent toujours rares, se sont développées.

Finalement, le même schéma se retrouve concernant l'homogamie d'origine sociale, mais avec un certain retard lié au retour en arrière d'une génération. Cette dimension a cependant pour particularité de mettre encore plus nettement en évidence que la précédente

la perte d'unité des classes populaires, ce que l'on peut attribuer aux effets de la massification scolaire et de la mobilité sociale.

Ces évolutions de l'homogamie absolue décrivent une transformation majeure de la structure de la société française. Qu'elles soient dues à la seule évolution de la composition de la population ou qu'elles résultent de surcroît d'une moindre détermination du choix du conjoint, elles ont eu des conséquences importantes sur les relations entre groupes sociaux, notamment en réduisant la cohésion des groupes d'indépendants et des classes populaires. Ce premier constat tend à valider la théorie de la modernisation, en particulier du point de vue de l'effet du déclin de la population agricole sur le niveau de l'endogamie : pays assez tardivement agricole, la France constitue, encore dans la période étudiée, un exemple assez net d'un tel processus. Il est très nettement apparu que l'endogamie des classes sociales en croissance est loin d'être assez forte pour compenser l'effet de la baisse de la proportion d'agriculteurs – et ce, même en considérant qu'employés et ouvriers forment une seule et même classe sociale. Dans le chapitre suivant, nous confirmerons que l'homogamie s'est affaiblie au-delà de ce que ces évolutions structurelles impliquaient, en systématisant les observations réalisées ci-dessus à partir des *odds ratios* généralisés.

CHAPITRE 2

L'HOMOGAMIE RELATIVE DE DIPLÔME, DE CLASSE ET D'ORIGINE SOCIALES ET SES ÉVOLUTIONS : OUVERTURE D'ENSEMBLE, REPLI DES ÉLITES

Après avoir étudié au chapitre précédent les évolutions de la structure de la population et de l'homogamie absolue, nous nous tournons vers l'analyse des variations de l'homogamie relative, c'est-à-dire des transformations du choix du conjoint au-delà de ce qu'impose la structure de la population. En effet, seule cette approche de l'homogamie permet d'évaluer dans quelle mesure le choix du conjoint parmi l'ensemble des partenaires possibles est socialement déterminé, reflétant l'effet des opportunités de rencontre, des préférences individuelles et des normes sociales. Elle autorise aussi une mesure de l'intensité d'ensemble de l'association entre caractéristiques des conjoints, sans s'en tenir aux seuls couples dans lesquels les conjoints appartiennent au même groupe (couples endogames).

Afin d'analyser en détail les évolutions de l'homogamie relative, nous recourons à des modèles log-linéaires et log-multiplicatifs capables de synthétiser les nombreux *odds ratios* qui peuvent être calculés dans les 42 tables d'homogamie annuelles construites à partir des enquêtes Emploi. Nous commençons par présenter le principe de la modélisation log-linéaire, ainsi que les critères et la démarche de sélection des modèles sur laquelle nous nous appuyerons dans ce chapitre et les suivants. Nous introduisons ensuite un indice de l'intensité de l'association, l'*odds ratio* absolu moyen (MAOR), que nous utiliserons pour mesurer la force de l'homogamie relative selon différentes composantes. Le corps des analyses est enfin articulé en deux étapes. Dans une première étape, nous mettons en évidence un affaiblissement de l'intensité générale de l'homogamie relative en termes de diplôme, de classe et d'origine sociale, avant de confirmer ce constat concernant

l'endogamie relative de la plupart des groupes – à l'exception notable des diplômés des grandes écoles. Dans une seconde étape, nous décomposons la structure de l'homogamie en différentes dimensions évoluant au cours du temps, considérant ainsi l'homogamie comme une manière de donner un contenu empirique à la notion d'espace social ; cette approche confirme l'affaiblissement de l'homogamie relative, tout en éclairant ses origines et en la nuanciant. En conclusion, nous tentons de mettre en relation nos résultats avec les travaux récents portant sur les classes sociales, la ségrégation et les inégalités, afin d'éclairer les nouvelles lignes de stratification de la société française.

1 Modéliser l'homogamie relative

L'étude de l'homogamie relative directement à partir des *odds ratios* ou des *odds ratios* généralisés présentée au chapitre précédent atteint rapidement ses limites lorsque l'on cherche à systématiser les analyses déjà présentées, dans l'objectif de mesurer l'évolution de l'homogamie relative à la fois en termes de *niveau* (intensité générale de l'homogamie) et de *structure* (proximités et distances entre groupes) au cours des 42 années d'enquête. Le nombre d'*odds ratios* à considérer rend en effet toute tentative de synthèse fort aventureuse sans l'aide d'une technique statistique permettant de faire émerger les traits majeurs de l'information disponible.

Du fait de la forte inertie du phénomène social considéré, les tables d'homogamie obtenues à partir de deux années successives sont loin d'être complètement différentes au point de justifier deux analyses séparées. Au contraire, les évolutions d'une année sur l'autre ont toutes les chances d'être très légères, et de ne devenir véritablement significatives qu'à plusieurs enquêtes d'intervalle. En outre, du fait de l'existence d'une composante aléatoire liée à l'utilisation d'une enquête par sondage, on peut s'attendre à ce qu'une partie non négligeable des différences entre deux années proches ne reflètent aucune évolution réelle de la population étudiée.

La modélisation log-linéaire (Bishop, Fienberg et Holland, [1975] 2007 ; Agresti, [1990] 2002) permet de répondre à ces difficultés. Un modèle log-linéaire n'est rien d'autre qu'une manière de synthétiser les *odds ratios* d'une table suivant différentes spécifications plus ou moins complexes, l'objectif étant de concilier simplicité (et qualité de l'interprétation) du

modèle et précision de la description des données¹. Parmi la très grande variété de modèles qui ont été utilisés dans la littérature, nous retiendrons plus bas ceux qui nous semblent les plus adaptés à nos données, c'est-à-dire des tables à trois dimensions (catégorie de l'homme, catégorie de la femme, année d'enquête). La troisième dimension, en tant que variable temporelle, a un caractère ordonné : elle peut être traitée de manière plus parcimonieuse qu'une variable non ordonnée. Ainsi, les modèles les plus adaptés décriront assez précisément l'association entre les deux premières dimensions (caractéristiques des conjoints), mais résumeront de manière beaucoup plus parcimonieuse les différences entre deux années (niveaux de la troisième dimension, dits généralement couches).

Trois modèles de référence méritent d'être introduits dès maintenant ; leurs équations sont détaillées dans l'encadré 2.1 (p. 136). Le premier est le modèle d'indépendance conditionnelle, qui suppose que l'homogamie relative est nulle chaque année d'enquête : les marges de la table (structure de la population) suffisent à reconstruire complètement les données observées. Ce modèle, irréaliste, est l'équivalent de la situation de mise en couple aléatoire introduite au chapitre précédent ; il sert de base pour mesurer l'apport des modèles plus élaborés à la description de l'homogamie. Le second modèle, sociologiquement plus crédible, est celui de stabilité de l'association. Partant du précédent, ce modèle introduit des paramètres permettant de décrire parfaitement l'association entre caractéristiques des conjoints, mais en supposant que celle-ci est restée identique pour toutes les années d'enquête au-delà des variations de la structure de la population. C'est ce modèle simple qui doit être rejeté puis dépassé si on observe effectivement une évolution de l'homogamie relative au cours du temps. Finalement, le troisième modèle de référence est le modèle saturé, qui décrit parfaitement les données observées. Ce modèle n'a pas d'intérêt en lui-même puisqu'il revient à interpréter directement les *odds ratios* de la table. Il permet cependant d'évaluer si, étant donné l'échantillon dont nous disposons, des modèles plus simples sont suffisants pour décrire les données aux erreurs d'échantillonnage près.

Ainsi, la procédure de sélection des modèles est au cœur de l'analyse des tables de contingence. Celle-ci repose à la fois sur des critères d'interprétation – certains modèles sont sociologiquement plus intéressants que d'autres – et sur des critères statistiques. Nous

1 Soulignons que les modèles log-linéaires et log-multiplicatifs ne possèdent pas la propriété essentielle d'indépendance des marges de la table s'ils ne s'ajustent pas correctement aux données (Breen, 2004, p. 405-407), d'où l'importance de la procédure de sélection des modèles développée plus loin.

décrivons dans la section suivante les difficultés que soulève cette procédure essentielle, avant de décrire les modèles que nous retiendrons ici.

1.1 La sélection des modèles : les indicateurs BIC et AIC

La question des critères de sélection d'un modèle fait l'objet de débats importants, à la hauteur de son caractère décisif pour l'analyse des tables de contingence – et plus largement pour toute démarche de modélisation. L'approche la plus classique consiste à comparer des modèles imbriqués (chacun comprenant le précédent comme cas particulier), en cherchant à savoir si les paramètres supplémentaires introduits dans le modèle le plus complexe sont justifiés par une amélioration de la description des données observées supérieure au bruit lié à l'échantillonnage aléatoire. Une telle procédure permet par exemple d'examiner si nos données permettent d'affirmer l'existence d'une homogamie relative en comparant le modèle d'association stable déjà décrit à celui d'indépendance conditionnelle. Ou encore de savoir si nous pouvons affirmer que l'homogamie relative a varié dans le temps, en comparant le modèle saturé au modèle de stabilité.

Cette comparaison, dénommée test du rapport de vraisemblance, est réalisée à partir de la déviance des modèles. Cet indicateur, qui mesure la qualité de l'ajustement d'un modèle aux données observées, est d'autant plus faible que le modèle en est proche. Notée G^2 ou L^2 , la déviance est définie comme :

$$G^2 = -2 \log V = 2 \sum_{k=1}^K n_k \log \frac{n_k}{m_k}$$

avec V la vraisemblance du modèle, K le nombre de cellules de la table, n_k les effectifs observés dans la cellule k , m_k ceux prédits par le modèle pour la même cellule. Par définition, la déviance du modèle saturé, qui décrit parfaitement les données, est nulle.

Le test du rapport de vraisemblance repose sur le fait que la différence entre les déviances de deux modèles suit asymptotiquement une loi du χ^2 avec un nombre de degrés de liberté égal au nombre de paramètres introduits dans le modèle le plus complexe par rapport au modèle plus simple. Il est ainsi possible de calculer une probabilité critique (*p-value*) indiquant s'il faut ou non, au regard des données disponibles, rejeter la restriction imposée par le choix du modèle le plus simple.

Cette méthode de comparaison présente cependant l'inconvénient de ne s'appliquer qu'à des modèles emboîtés, alors qu'il est fréquent de vouloir tester l'ajustement de modèles qui diffèrent sous plusieurs aspects, sans que l'un soit un cas particulier de l'autre. C'est l'une des raisons pour lesquelles d'autres indicateurs statistiques sont souvent utilisés. Les plus populaires sont le critère d'information bayésien (BIC, aussi appelé critère de Schwarz, voir Schwarz, 1978 ; Raftery, 1986, 1995) et le critère d'information d'Akaike (AIC, voir Akaike, 1974). Tous deux sont fondés, comme le test du rapport de vraisemblance, sur la déviance des modèles et leur nombre de degrés de liberté. Ils sont définis comme suit :

$$AIC = G^2 - 2d$$

$$BIC = G^2 - d \log n$$

avec d le nombre de degrés de liberté (soit le nombre de cellules K moins le nombre de paramètres indépendants du modèle), et n la taille de l'échantillon. On voit donc que la diminution de la déviance due à la plus grande complexité d'un modèle est mise en balance avec le nombre de degrés de liberté que cette amélioration a coûté. Les deux indicateurs sont égaux à zéro pour le modèle saturé, qui par définition présente une déviance nulle et aucun degré de liberté. Une valeur négative indique donc qu'un modèle doit être préféré au modèle saturé ; plus généralement, le modèle ayant la valeur la plus faible (ou la plus négative) doit être retenu : il présente le meilleur rapport entre qualité d'ajustement aux données et complexité de la spécification.

Contrairement au test du rapport de vraisemblance, il n'est pas nécessaire que des modèles soient emboîtés pour les comparer, ce qui fait du BIC et de l'AIC des indicateurs très flexibles pour la sélection du meilleur modèle. Mais ces critères ne résolvent pas pour autant toutes les difficultés.

1.2 Les limites de la sélection d'un modèle à partir d'un unique critère statistique

Dans la tradition de l'analyse de tables de contingence, c'est le plus souvent le BIC qui est utilisé pour comparer la qualité d'ajustement des différents modèles envisagés. Cette préférence pour le BIC plutôt que pour d'autres indicateurs, dont notamment l'AIC, ne paraît malheureusement pas véritablement reposer sur des fondements statistiques solides. Elle semble bien plus devoir, en termes purement pratiques, à une caractéristique bien

connue du BIC : celle de pénaliser plus fortement que le test du rapport de vraisemblance, et que l'AIC, le nombre de degrés de liberté d'un modèle, et donc de préférer des modèles plus simples². Cette propriété est souvent considérée comme souhaitable, alors qu'à l'inverse, le fait que les tests de significativité soient sensibles à la taille de l'échantillon utilisé est présenté comme une nuisance. En particulier, avec des échantillons de la taille de ceux typiquement utilisés pour les analyses de l'homogamie ou de la mobilité sociale, tous les modèles, aussi complexes soient-ils, sont rejetés par un test du rapport de vraisemblance, ne laissant apparemment au chercheur que la possibilité d'utiliser le modèle saturé, « vrai » par définition, mais comprenant trop de paramètres pour être interprétable.

Contre cette utilisation courante du BIC, et suivant en cela Andrew Gelman et Donald Rubin (1995), nous défendons ici la position selon laquelle il est illusoire d'attendre d'un indicateur statistique *omnibus* qu'il réponde automatiquement à des questions scientifiques de natures très diverses. Alors qu'aucun sociologue ne prendrait au sérieux l'idée que l'un des modèles simplifiés testés puisse épuiser complètement la réalité sociale qu'il tente de résumer, comment pourrait-on attendre d'un critère qu'il sélectionne automatiquement un modèle autre que le modèle saturé lorsque l'échantillon est de taille importante ? Comment un critère statistique pourrait-il choisir un modèle qui ne s'ajuste pas correctement aux données en le présentant comme plus probable que le modèle saturé (Gelman et Rubin, 1995, p. 168) ? L'affirmation selon laquelle les tests classiques sont trop sensibles aux effectifs en jeu repose *in fine* sur le postulat, à notre sens erroné, selon lequel il est possible de trouver un critère automatique qui permettrait de répondre à toutes les questions scientifiques sans faire intervenir le jugement du chercheur. En effet, il est pour le moins paradoxal de regretter que tous les tests statistiques soient significatifs sur un échantillon, alors que ces tests devraient constituer une contrainte dont la levée ne peut être qu'une bonne nouvelle pour le chercheur : dans la réalité – c'est-à-dire sur population entière, par exemple sur des données de recensement – *tous les effets sont significatifs* (Gelman et Rubin, 1999, p. 408). Si un échantillon est assez grand pour autoriser une très grande finesse de description, c'est sur des critères scientifiques concernant le caractère négligeable de telle différence d'ordre secondaire, ou au contraire l'importance sociologique d'un effet qui aurait pu passer pour

2 Dans la définition du BIC donnée plus haut, cette propriété est visible au fait qu'un modèle est pénalisé pour chaque degré de liberté à proportion du logarithme des effectifs de la table.

secondaire si la théorie ne lui accordait un statut particulier, que doit s'appuyer l'analyse (Gelman et Rubin, 1995 ; Firth et Kuha, 1999).

Il s'agit ainsi de dépasser la question de la significativité statistique pour passer à celle de la significativité sociologique ou substantive (Selvin, 1957). Le recours inconditionnel au BIC apparaît comme une substitution abusive de la première à la seconde, alors que les considérations théoriques devraient toujours avoir le dernier mot. En effet, suivant la fameuse citation de George Box (voir notamment Box et Draper, 1987, p. 424 ; Burnham et Anderson, 2004), « *tous les modèles sont faux, mais certains sont utiles* » : lorsque l'échantillon est assez important pour rejeter tous les modèles comme faux, il s'agit pour le sociologue de juger lesquels sont les plus utiles.

Nous ne considérons pas pour autant que les critères d'information statistique sont inutiles. Leur rôle est essentiel pour mettre à l'épreuve différentes manières de résumer le réel et tester des hypothèses concernant la structure de l'association observée. Mais le chercheur doit toujours rester libre de s'en tenir à un modèle plus simple si son interprétation permet de faire émerger des interprétations. Cette démarche nous paraît valide tant qu'il demeure tout à fait clair que ce modèle est une simplification de la réalité, et que l'ampleur de cette simplification est mesurée de manière explicite (Erikson et Goldthorpe, 1987, p. 60). Ainsi, une démarche scientifique rigoureuse peut consister, non pas à adopter systématiquement le modèle préféré par un critère, mais à partir d'un modèle dont l'interprétation semble fructueuse, et à le compléter par des paramètres reflétant son écart au modèle saturé, ou à un modèle mieux ajusté aux données (Gelman et Rubin, 1995). On est ainsi en position de mesurer l'adéquation aux données de l'hypothèse que constitue le modèle, aussi bien que ses limites.

1.3 Les objectifs différents du BIC et de l'AIC

Outre ces remarques générales concernant le recours à un indicateur de qualité du modèle, quel qu'il soit, d'importants débats ont porté sur les mérites respectifs des différents critères, BIC et AIC en tête. Les critiques du BIC ont été particulièrement fortes, alors que l'AIC, qui se révèle – comme on va le voir – plus proche de la théorie classique de la vraisemblance, semble avoir été moins attaqué sur le plan théorique.

Les critiques du BIC sont relativement anciennes en sociologie : en fait, elles apparaissent presque aussi anciennes que l'utilisation du BIC lui-même. Robert Erikson et John Goldthorpe (1992, p. 101, 326) soulignaient déjà l'usage incohérent de cet indicateur dans la littérature portant sur la mobilité sociale, et pointaient le fait que les modèles qui avaient réellement la faveur du BIC étaient souvent ignorés par les auteurs à cause de leur faible crédibilité sociologique. Dans cette logique, certains travaux récents (Halpin et Chan, 2003 ; Breen, 2004) ne recourent ainsi qu'au test du rapport de vraisemblance.

La critique la plus développée du BIC, qui semble avoir fait date dans la discipline, a été formulée par David Weakliem (1999) dans un article qui, suite au débat suscité au sein du comité de lecture, a donné lieu à un numéro spécial de la revue *Sociological Methods & Research*. L'auteur y soulignait notamment deux faiblesses ou incohérences de l'indicateur. En premier lieu, il montrait que, malgré sa prétention à la neutralité, le BIC repose sur le choix d'une distribution des probabilités *a priori* des paramètres très spécifique, d'autres pouvant trouver des justifications tout aussi solides ; cette distribution présente en outre la particularité d'être dépendante de la taille de l'échantillon utilisé. De manière plus problématique, l'auteur mettait en évidence le fait que la taille de l'échantillon, qui entre dans la formule de l'indicateur, n'est pas une bonne mesure de la quantité d'information disponible : à taille d'échantillon égale, la distribution des individus influe fortement sur la possibilité d'étudier un effet donné. À titre d'exemple, dans une comparaison de moyennes entre deux groupes de tailles très inégales, agrandir l'échantillon en ajoutant un nombre donné d'individus au groupe le plus grand n'augmente pas autant la capacité à mesurer l'écart entre les groupes que d'ajouter ces individus au groupe le plus petit. Cette particularité du BIC donne parfois lieu à des résultats peu cohérents : dans une comparaison de la mobilité sociale entre pays, inclure les pays avec les plus faibles échantillons amène à préférer un modèle plus simple au modèle saturé, alors que l'inverse est vrai si ces pays sont exclus, quand bien même le modèle simple décrit moins bien les données dans ces pays que dans les autres. En d'autres termes, l'analyse conjointe de tous les pays n'est pas, aux yeux du BIC, équivalente à la combinaison des analyses pays par pays (ce qui est au contraire le cas avec le test du rapport de vraisemblance ou l'AIC). De même, le BIC donne des résultats différents selon que l'on insère des paramètres spécifiques à certaines cellules ou qu'on les ignore lors de l'estimation du modèle, alors que les paramètres estimés restent les mêmes.

Ces critiques paraissent défendre le recours à l'AIC plutôt qu'au BIC comme critère de sélection³. En effet, celui-ci tient compte de la taille de l'échantillon par le biais de la vraisemblance du modèle, mais pas de manière extérieure à celle-ci : il n'est donc pas sujet aux difficultés tenant à la définition de la taille effective de l'échantillon. Cette propriété peut être reliée au fait (Burnham et Anderson, 2004) que l'AIC constitue un estimateur asymptotiquement sans biais de la divergence de Kullback-Leibler (ou entropie relative) du modèle considéré par rapport à un modèle de référence (par exemple le modèle saturé, mais potentiellement n'importe quel modèle). Ce lien direct entre l'AIC et une mesure fondamentale de la théorie de l'information, qui constitue une extension de la théorie de la vraisemblance, lui fournit une solide justification théorique : cet indicateur mesure la perte d'information due à l'utilisation d'un modèle par rapport à un autre. De fait, dans le cas particulier des modèles emboîtés, réaliser un test du rapport de vraisemblance en se dotant d'un seuil de significativité est strictement équivalent à vérifier que la différence d'AIC entre les modèles est supérieure à une certaine valeur (Murtaugh, 2014). L'utilisation de l'AIC permet de généraliser le principe de ce test à la comparaison de modèles non nécessairement emboîtés, et ainsi d'évaluer simultanément la vraisemblance de plusieurs hypothèses (Burnham et Anderson, 2014).

Mais la différence entre BIC et AIC tient surtout aux objectifs qu'ils visent. Montrant que les deux indicateurs peuvent se dériver dans un cadre bayésien, en recourant à des distributions de probabilité *a priori* très différentes, Kenneth Burnham et David Anderson (2004) ont mis en évidence la différence de philosophie entre BIC et AIC (voir aussi Aho, Derryberry et Peterson, 2014). Le premier cherche ainsi à identifier un modèle dit quasi-vrai, dont la probabilité par rapport aux autres modèles testés tend vers l'unité lorsque la taille de l'échantillon augmente, et ce, quelle que soit la taille de l'échantillon. Au contraire, l'AIC considère que le modèle ciblé change lorsque l'échantillon grandit : plus l'échantillon est important, et plus le modèle pourra être complexe. Ces différences d'approche impliquent que BIC est plus efficace lorsque le vrai modèle contient quelques effets très forts. Au contraire, l'AIC est à préférer s'il existe une série d'effets dont l'importance décroît de manière régulière. Le BIC aura dans ce dernier cas tendance à être trop parcimonieux, alors

3 Alan Agresti ([1990] 2002) ne mentionne par exemple le BIC qu'en note, renvoyant aux débats que nous venons de citer, et présente l'AIC comme l'indicateur le plus solide. Ronald Christensen (1994) ne mentionne et recommande que l'AIC.

qu'à l'inverse l'AIC pourra choisir dans le premier cas un modèle trop complexe. Ainsi, le BIC semble être plutôt adapté aux cas où la réalité suit un modèle relativement simple, quand l'AIC est adapté à ceux où le modèle générateur des données est très complexe et inconnu ; ou encore, le premier dans le cadre d'analyses confirmatoires d'hypothèses clairement établies, le second d'analyses exploratoires ou de prédictions (Aho, Derryberry et Peterson, 2014).

Il peut donc sembler paradoxal que l'usage du BIC soit si répandu dans les sciences sociales, domaine où la notion de « vrai modèle » semble – au mieux – fort abstraite. Concernant plus spécifiquement l'analyse des tables de contingence, dans la mesure où les modèles testés constituent généralement des améliorations progressives du modèle d'indépendance jusqu'au modèle saturé, il paraît difficile de considérer qu'il existe un modèle générateur simple. Au contraire, les effets pris en compte diminuent généralement en amplitude : les modèles les plus simples décrivent déjà l'essentiel des données, les modèles suivants n'apportant que des améliorations d'importance décroissante. Dans une perspective temporelle ou comparative, le résultat majeur est généralement constitué par la relative stabilité de la structure de l'association (Erikson et Goldthorpe, 1992), alors que des déviations moins importantes peuvent être notées au cours du temps et entre pays. Est-ce à dire que le « vrai modèle » est celui d'une invariance temporelle et spatiale ? Il nous semble que le BIC conduit trop souvent à conclure dans ce sens, là où l'AIC ou le test du rapport de vraisemblance révéleraient l'existence de différences, certes secondaires, mais parfois d'un grand intérêt scientifique – ce qu'aucun critère ne peut deviner seul.

Au total, si les indicateurs BIC et AIC nous paraissent constituer un guide précieux pour l'évaluation de la qualité des différents modèles, ils ne constituent qu'un élément parmi d'autres. Nous ferons donc usage simultanément de plusieurs indicateurs (Xie, 1999, p. 433), sans pour autant nous contenter de constater leurs possibles divergences : au vu des éléments présentés ci-dessus, nous savons plus précisément que le BIC aura plutôt tendance à choisir des modèles trop simples, alors que l'AIC défendra le choix de modèles plus élaborés. Il s'agira donc d'évaluer selon d'autres critères si le modèle choisi par le BIC est assez complexe pour les questions que nous nous posons, ou s'il est nécessaire d'aller au-delà⁴.

4 Soulignons en outre que les indicateurs BIC et AIC n'ont que très récemment été adaptés à l'analyse de données issues d'enquêtes complexes (Lumley et Scott, 2015) comme les enquêtes Emploi (*cf.* chapitre

1.4 Des indicateurs absolus d'ajustement

Cette démarche est rendue possible par l'existence de critères statistiques permettant de mesurer la qualité *absolue* des modèles, alors que les indicateurs comme le BIC et l'AIC, n'autorisent que l'évaluation de leur qualité relative (comparaison entre modèles). Les indicateurs de qualité absolue sont nécessaires pour vérifier que le modèle retenu décrit de manière acceptable les données, de manière à mesurer la validité des interprétations qu'il autorise. En effet, si aucun modèle correct n'est testé (en plus du modèle saturé), le BIC et l'AIC choisiront le moins mauvais modèle, qui peut demeurer assez fortement éloigné des données.

Deux indices absolus sont habituellement utilisés pour évaluer la qualité d'ajustement d'un modèle log-linéaire ou log-multiplicatif : l'indice de dissimilarité, et le pourcentage de la déviance expliquée.

L'indice de dissimilarité (*cf.* Kuha et Firth, 2011 pour une analyse approfondie récente) offre l'avantage d'une mesure intuitive de la qualité de la prédiction d'un modèle. Développé originellement pour la mesure de la ségrégation spatiale, il correspond dans une analyse log-linéaire à la proportion des effectifs qui sont mal décrits par le modèle – ou plus précisément, à ceux qui devraient être déplacés d'une cellule vers une autre pour que la table de contingence prédite par le modèle corresponde exactement à la table observée. Formellement, cet indice, noté Δ , est défini comme :

$$\Delta = \frac{1}{2N} \sum_{k=1}^K |n_k - m_k|$$

avec N la taille de l'échantillon, K le nombre total de cellules, n_k l'effectif observé dans la cellule k , et m_k celui prédit par le modèle dans la même cellule.

Par définition, le modèle saturé, reproduisant parfaitement les données, présente une dissimilarité de 0. Une valeur de 1 (ou 100 %) ne peut jamais être atteinte, même par un modèle très éloigné des données. La qualité absolue d'un modèle se juge donc à sa proximité d'un côté au modèle saturé, de l'autre au modèle d'indépendance (postulant qu'aucune

précédent). La théorie permettant l'application de ces indicateurs modifiés faisant encore défaut pour les modèles log-multiplicatifs que nous utilisons, nous recourons aux indicateurs BIC et AIC standard, qui reposent sur l'hypothèse que les données sont issues d'un tirage aléatoire simple, et qui auront pour cela tendance à choisir des modèles plus complexes que nécessaire. Cela ne pose pas de problème majeur ici : les différences entre modèles sont trop flagrantes pour résulter de ce seul biais.

association n'existe entre les variables étudiées au-delà de l'effet structurel lié à la répartition de la population entre les différentes catégories). Plus généralement, l'indice permet aussi d'évaluer la qualité relative d'un modèle en le comparant à n'importe quel autre.

En dépit de l'interprétation assez intuitive de cet indicateur, soulignons qu'il demeure relativement difficile de définir ce qui constitue une différence importante ou au contraire faible en termes de dissimilarité. Nous retrouvons les critiques adressées dans la section précédente à l'utilisation aveugle d'un indicateur, quel qu'il soit : une différence faible peut mériter notre attention si elle permet d'éclairer une question théorique importante – par exemple celle de l'hypergamie féminine. Néanmoins, en donnant la mesure de la proportion des effectifs en jeu dans la différence entre deux modèles, l'indice de dissimilarité permet de ramener la comparaison entre deux modèles – souvent abstraite lorsque fondée uniquement sur le BIC ou l'AIC – à sa juste mesure au regard de la population totale. Nous en verrons bien des exemples plus bas.

Dans une logique similaire, mais en se fondant cette fois sur la déviance, il est possible de mesurer la qualité du modèle en le comparant d'une part au modèle saturé (dont la déviance est nulle), et d'autre part au modèle d'indépendance. La part de la déviance du modèle d'indépendance expliquée par un modèle donné, parfois notée rG^2 , constitue ainsi un indicateur intéressant du progrès réalisé dans l'explication du phénomène étudié. Par rapport à l'indice de dissimilarité, la part de déviance expliquée présente l'avantage de pondérer les écarts par les effectifs réellement observés – principe qui est au fondement de l'estimation par le maximum de vraisemblance. En effet, un écart d'un individu n'a pas la même signification dans une cellule vide et dans une cellule à très fort effectif. En revanche, l'inconvénient de ce critère est de ne pas offrir d'interprétation directe en termes du nombre d'individus concernés. Pour cette raison, nous lui préférons généralement par simplicité l'indice de dissimilarité.

1.5 La stratégie de modélisation

De manière à mesurer à la fois des tendances générales et des transformations de la structure sociale, nous adoptons une double stratégie de modélisation. Dans un premier temps, le modèle UNIDIFF (ou modèle log-multiplicatif à effet de couche, Erikson et

Goldthorpe, 1992 ; Xie, 1992) nous permettra d'observer les variations temporelles du niveau de l'homogamie relative, en faisant l'hypothèse d'une stabilité de la structure de l'association entre les caractéristiques des conjoints, c'est-à-dire la stabilité de tous les *odds ratios* de la table, à l'exception d'un coefficient annuel qui les affecte tous de manière identique.

Afin de préciser l'analyse en étudiant séparément l'évolution de l'endogamie de chaque groupe, nous recourons ensuite au modèle log-multiplicatif à forme de régression (Goodman et Hout, 1998). Ce modèle postule que l'homogamie a varié sur la période étudiée entre un état de départ et un état d'arrivée, des coefficients annuels mesurant la position dans la transition entre les deux. Du fait de la taille importante de nos tables, nous n'interprétons pas les *odds ratios* correspondant à chacune des configurations de couple, mais seulement ceux concernant les configurations endogames (diagonale de la table).

Dans un second temps, nous décrivons les configurations exogames (cellules hors diagonale⁵) en utilisant des modèles d'association lignes colonnes avec effet de couche RC-L. De manière similaire à une analyse des correspondances dont l'importance des dimensions varierait au cours du temps, ces modèles permettent de résumer de manière visuelle la structure de l'espace social et son évolution temporelle. Bien qu'ils soient à notre sens d'un grand intérêt, ces modèles puissants ont rarement été utilisés pour l'étude de l'homogamie – et seulement dans leur version à une dimension (Wong, 2003 ; Domański et Przybysz, 2007). Nous complétons ce tableau par des représentations de l'espace social à une date donnée à l'aide de modèles RC (c'est-à-dire sans évolution temporelle), afin de donner une image de la signification des transformations temporelles mises en évidence.

Puisque nos données sont constituées d'un nombre relativement important de points temporels, nous testons systématiquement la présence d'une tendance linéaire. Cependant, nous représentons toujours dans les figures les coefficients annuels (non linéaires) étant donné qu'ils permettent mieux d'évaluer les tendances ainsi que l'homogénéité des séries d'enquêtes, ce qui constitue l'un des intérêts du recours à des enquêtes annuelles.

5 Pour cela, les cellules de la diagonale sont exclues de la modélisation. Ceci revient à leur attribuer des poids nuls tout en ajustant le nombre de degrés de liberté du modèle, ou encore à décrire parfaitement la diagonale avec un paramètre par cellule pour chaque année. Les indicateurs d'ajustement des modèles que nous présentons ignorent les cellules de la diagonale : ils ne sont pas artificiellement améliorés du fait de la description parfaite de ces cellules.

Nous utilisons des quasi-intervalles de confiance (Firth et Menezes, 2004), plutôt que des intervalles de confiance classiques. En effet, ces derniers ne permettent de comparer un coefficient qu'avec la référence, fixée par convention à 0 ou 1 selon les cas : les comparaisons entre deux années ne sont pas rigoureusement possibles. Les quasi-intervalles de confiance résolvent ce problème en offrant une approximation des intervalles de confiance qui seraient obtenus en prenant successivement chaque coefficient pour référence.

Tous les modèles sont estimés dans l'environnement R (R Core Team, 2014) en utilisant les paquets *gnm* (Turner et Firth, 2014), *survey* (Lumley, 2014) et *logmult* (Bouchet-Valat, 2015a)⁶.

Encadré 2.1 : Équations des modèles mentionnés

Tous les modèles mentionnés prennent pour base le modèle d'indépendance conditionnelle. En notant m_{hft} les effectifs prédits par le modèle pour la cellule à l'intersection de la ligne h (catégorie de l'homme), de la colonne f (catégorie de la femme) et de la couche t (année d'enquête), dans une table de dimensions $H \times F \times T$, ce modèle s'écrit :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT}$$

Ce modèle présente $(H-1)(F-1)T$ degrés de liberté. Tous les *odds ratios* conditionnels contrastant deux catégories masculines à deux catégories féminines une année donnée sont égaux à 1 :

$$\theta_{hf, h'f'(t)} = 1$$

Le modèle d'association stable ajoute à ce modèle l'interaction entre caractéristiques des conjoints :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + \lambda_{hf}^{HF}$$

Ce modèle présente $(H-1)(F-1)(T-1)$ degrés de liberté. Les *odds ratios* conditionnels sont les mêmes pour toutes les années d'enquête :

⁶ Merci à Heather Turner pour ses réponses aux nombreuses questions posées durant le développement du paquet *logmult*, ainsi que pour les améliorations qu'elle a apportées au paquet *gnm* suite à nos remarques.

$$\theta_{hf,h'f'(t)} = \exp \left[\lambda_{hf}^{HF} + \lambda_{h'f'}^{HF} - \lambda_{hf'}^{HF} - \lambda_{h'f}^{HF} \right]$$

Le modèle saturé, reproduisant parfaitement les données, ajoute à ce dernier une interaction (de second ordre) entre caractéristiques des conjoints et année d'enquête :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + \lambda_{hf}^{HF} + \lambda_{hft}^{HFT}$$

Ce modèle ne présente aucun degré de liberté. Les *odds ratios* conditionnels diffèrent complètement d'une année d'enquête à l'autre :

$$\theta_{hf,h'f'(t)} = \exp \left[\left(\lambda_{hf}^{HF} + \lambda_{h'f'}^{HF} - \lambda_{hf'}^{HF} - \lambda_{h'f}^{HF} \right) + \left(\lambda_{hft}^{HFT} + \lambda_{h'f't}^{HFT} - \lambda_{hf't}^{HFT} - \lambda_{h'f't}^{HFT} \right) \right]$$

Les autres modèles utilisés se situent entre le modèle d'indépendance conditionnelle et le modèle saturé : ils proposent des spécifications plus parcimonieuses des coefficients d'interaction λ_{hf}^{HF} et λ_{hft}^{HFT} . Le modèle log-multiplicatif avec effet de couche ou UNIDIFF (Erikson et Goldthorpe, 1992 ; Xie, 1992) enrichit celui d'association stable en autorisant l'interaction entre caractéristiques des conjoints à varier de manière uniforme selon l'année, suivant l'équation :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + \phi_t \psi_{hf}$$

Ce modèle présente $(HF-H-F)(T-1)$ degrés de liberté. Les *odds ratios* conditionnels sont proportionnels sur l'échelle logarithmique entre années d'enquête :

$$\theta_{hf,h'f'(t)} = \exp \left[\phi_t \left(\psi_{hf} + \psi_{h'f'} - \psi_{hf'} - \psi_{h'f} \right) \right]$$

Le modèle à forme de régression (Goodman et Hout, 1998) complexifie encore le modèle UNIDIFF en postulant l'existence d'une partie stable de l'association en plus de la partie variable au cours du temps :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + \lambda_{hf}^{HF} + \phi_t \psi_{hf}$$

Ce modèle présente $(HF-H-F)(T-2)$ degrés de liberté. Les *odds ratios* conditionnels sont liés par une fonction affine sur l'échelle logarithmique entre années d'enquête :

$$\theta_{hf,h'f'(t)} = \exp \left[\left(\lambda_{hf}^{HF} + \lambda_{h'f'}^{HF} - \lambda_{hf'}^{HF} - \lambda_{h'f}^{HF} \right) + \phi_t \left(\psi_{hf} + \psi_{h'f'} - \psi_{hf'} - \psi_{h'f} \right) \right]$$

Le modèle d'association lignes-colonnes avec effet de couche ou RC(M)-L enrichit le modèle d'association conditionnelle en suivant une autre approche que les deux précédents. Dans la version retenue ici, les scores sont supposés stables dans le temps, alors que les coefficients d'association intrinsèque sont autorisés à varier :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + \sum_{m=1}^M \phi_{mt} \mu_{hm} v_{fm}$$

M est le nombre de dimensions du modèle, ϕ_{mt} le coefficient d'association intrinsèque relatif à la dimension m pour l'année t , μ_{hm} et v_{fm} les scores correspondant respectivement aux catégories des hommes et des femmes pour la dimension m . Les scores sont rendus identifiables en leur imposant par convention d'être centrés et réduits, ce qui est ici effectué en appliquant une pondération par la moyenne des marges en lignes et en colonnes de la table sur l'ensemble des années :

$$\sum_{h=1}^H \mu_{hm} P_{h++} = \sum_{f=1}^F v_{fm} P_{+f+} = 0$$

$$\sum_{h=1}^H \mu_{hm}^2 P_{h++} = \sum_{f=1}^F v_{fm}^2 P_{+f+} = 1$$

Ce modèle présente $(H-1)(F-1)T - M(H+F+T-4)$ degrés de liberté : il peut donc être plus ou moins parcimonieux qu'UNIDIFF et que le modèle à forme de régression selon la taille de la table et le nombre de dimensions introduites. Les *odds ratios* conditionnels se décomposent en un produit de contributions par dimension, qui sont elles-mêmes fonction de l'intensité de chaque dimension et des écarts entre les scores des hommes et des femmes :

$$\theta_{hf,h'f'(t)} = \prod_{m=1}^M \exp \left[\phi_{mt} (\mu_{hm} - \mu_{h'm}) (v_{fm} - v_{f'm}) \right]$$

Ainsi, si la contribution de chaque dimension aux *odds ratios* conditionnels est proportionnelle sur l'échelle logarithmique entre années d'enquête, les *odds ratios*

conditionnels sont liés par une relation assez complexe du fait de la combinaison de plusieurs dimensions.

Le modèle RC(M) diffère des modèles précédents puisqu'il est appliqué à des tables à deux dimensions. En notant m_{hf} les effectifs prédits par le modèle pour la cellule à l'intersection de la ligne h (catégorie de l'homme), de la colonne f (catégorie de la femme) dans une table de dimensions $H \times F$, il s'écrit :

$$\log m_{hf} = \lambda + \lambda_h^H + \lambda_f^F + \sum_{m=1}^M \phi_m \mu_{hm} \nu_{fm}$$

De manière analogue au modèle RC(M)-L, M est le nombre de dimensions du modèle, ϕ_m le coefficient d'association intrinsèque relatif à la dimension m , μ_{hm} et ν_{fm} les scores correspondant respectivement aux catégories des hommes et des femmes pour la dimension m . Les scores sont rendus identifiables en leur imposant par convention d'être centrés et réduits, et d'être non corrélés entre dimensions, ce qui est ici effectué en appliquant une pondération par les marges en lignes et en colonnes de la table :

$$\begin{aligned} \sum_{h=1}^H \mu_{hm} P_{h+} &= \sum_{f=1}^F \nu_{fm} P_{+f} = 0 \\ \sum_{h=1}^H \mu_{hm}^2 P_{h+} &= \sum_{f=1}^F \nu_{fm}^2 P_{+f} = 1 \\ \sum_{h=1}^H \mu_{hm} \mu_{hm'} P_{h+} &= \sum_{f=1}^F \nu_{fm} \nu_{fm'} P_{+f} = 0 \text{ pour tout } m \neq m' \end{aligned}$$

Ce modèle présente $(H-M-1)(F-M-1)$ degrés de liberté. Les *odds ratios* se décomposent en un produit de contributions par dimension, qui sont elles-mêmes fonction de l'intensité de chaque dimension et des écarts entre les scores des hommes et des femmes :

$$\theta_{hf, h'f'} = \prod_{m=1}^M \exp \left[\phi_m (\mu_{hm} - \mu_{h'm}) (\nu_{fm} - \nu_{f'm}) \right]$$

Les *odds ratios* généralisés, qui sont ici conditionnels à une année d'enquête, peuvent être calculés directement à partir des coefficients log-linéaires d'interaction (Kaufman et Schervish, 1987). Sous le modèle de stabilité de l'association, ils sont simplement définis à partir du coefficient d'interaction correspondant à la cellule considérée :

$$GOR_{hf(t)} = \exp \left[\frac{HF}{(H-1)(F-1)} \lambda_{hf}^{HF} \right]$$

Cette formulation peut aisément être étendue aux modèles présentés plus haut, dont notamment le modèle à forme de régression :

$$GOR_{hf(t)} = \exp \left[\frac{HF}{(H-1)(F-1)} \left(\lambda_{hf}^{HF} + \phi_t \psi_{hf} \right) \right]$$

2 Mesurer l'intensité de l'homogamie relative

La littérature portant sur les modèles log-linéaires et log-multiplicatifs a développé de nombreuses spécifications permettant d'analyser la structure et le niveau de l'association, ainsi que ses variations au cours du temps, dans l'espace ou entre groupes. Néanmoins, de manière surprenante, aucun indice général de l'intensité de l'association – c'est-à-dire, en ce qui nous concerne, de la force de l'homogamie relative – n'a été proposé. Ainsi, lorsque les études de la mobilité sociale comparent différents pays du point de vue de l'intensité de la mobilité relative (Erikson et Goldthorpe, 1992 ; Breen, 2004), elles se fondent notamment sur le modèle UNIDIFF pour situer les pays sur une échelle les uns par rapport aux autres – mais aucune mesure *absolue* de l'intensité de l'association relative entre position d'origine et position occupée par les individus n'est fournie. Si elle ne soulève aucune difficulté pour les comparaisons conduites à l'intérieur d'un ouvrage, cette situation présente l'inconvénient d'interdire toute méta-analyse *a posteriori* visant à réaliser la synthèse de multiples travaux réalisés séparément, alors même que les méthodologies et les classifications mobilisées sont souvent assez proches. Ce défaut est particulièrement criant dans le champ de l'étude de l'homogamie, offrant beaucoup plus d'analyses portant sur une société en particulier, que de travaux de grande envergure réunissant de nombreux pays.

2.1 Les limites des indices existants

Sans doute ne faut-il pas exagérer l'ampleur de ce problème. En effet, certains modèles fournissent une mesure de l'intensité de l'association ; mais ces indices souffrent à nos yeux de plusieurs défauts. Ainsi, les modèles d'association lignes-colonnes – association uniforme, RC(M) ou encore RC(M)-L –, assez fréquemment utilisés dans la littérature (Smits, Ultee et Lammers, 1999 ; Wong, 2003 ; Katrňák, Fučík et Luijckx, 2012 ; Wong, 2003) offrent un indice de la force de l'association par le biais du coefficient d'association intrinsèque associé à chaque dimension. Néanmoins, dans leur version à une seule dimension, presque toujours retenue par les travaux existants, ces modèles reposent sur l'hypothèse que l'association peut être parfaitement décrite par une échelle unique – approximation qui nous semble trop restrictive, comme on le verra plus bas.

D'autres travaux recourent à des modèles de distance, dans lesquels un paramètre de pas (*step*) mesure la difficulté à former des couples suivant le nombre de catégories d'écart entre les conjoints (Smits, Ultee et Lammers, 1998, 2000 ; Domański et Przybysz, 2007). Une telle mesure offre une indication absolue de l'intensité de l'homogamie relative. Néanmoins, elle repose elle aussi sur une hypothèse forte, celle que le franchissement d'une catégorie a la même valeur dans tout l'espace social. En outre, on voit la difficulté qu'il y a à comparer les résultats des deux types assez différents de modèles que nous venons de présenter.

De manière plus problématique à nos yeux, ces deux approches tiennent mal compte des variations de l'association correspondant aux configurations endogames. En effet, il est fréquent que des paramètres spécifiques à la diagonale de la table soient introduits du fait de l'amélioration de l'ajustement du modèle qu'ils permettent. Dans ce cas, on ne dispose plus d'un indice unique résumant l'intensité de l'association endogame *et* exogame, ce qui amène par exemple Henryk Domański et Dariusz Przybysz (2007) à étudier les déterminants de la seconde uniquement.

À l'inverse, au moins deux publications ont proposé des indices se concentrant sur la mesure de l'intensité de l'endogamie relative. Ces deux indices suivent des logiques très similaires, et peuvent être utilisés en dehors de toute démarche de modélisation. Le premier, l'indice de diagonalité nette, a été proposé par Michel Forsé et Louis Chauvel (1995). Le second, l'indice H (pour indice d'homogamie), a été élaboré par Gunn Elisabeth Birkelund et Johan Heldal (2003). Fondés sur les *odds ratios* ou les coefficients d'interaction du modèle

log-linéaire saturé, ces indices contrastent l'association observée sur la diagonale de la table à celle observée dans le reste de la table⁷. L'attention particulière accordée par ces deux indices à la diagonale, correspondant aux couples endogames, nous paraît inutilement restrictive. En effet, cette approche revient à se doter d'une définition *a priori* de l'homogamie, contrainte à suivre une opposition binaire, plutôt que d'observer dans les faits quelles sont les attirances et répulsions entre les différents groupes sociaux et quelle est leur ampleur. Pour illustrer cette difficulté à partir d'un cas typique, un couple ouvrier-employée n'est pas homogame suivant la logique de ces deux indices, et la sur-représentation de ces configurations entraîne à leur sens une diminution de l'intensité de l'homogamie. Ainsi, ces indices ne nous semblent pas satisfaire à notre objectif, énoncé au chapitre précédent, de considérer l'homogamie comme un continuum en tenant compte de toute la diversité des configurations de couple observées.

2.2 Le coefficient d'association intrinsèque λ de Goodman

Un indice tenant compte des structures complexes que peut prendre l'association entre caractéristiques des conjoints nous semble ainsi nécessaire. Un tel indice a en fait déjà été proposé par Leo Goodman (1996), quoiqu'il soit resté quelque peu inaperçu à l'intérieur d'un article au caractère fortement théorique. Présentant un cadre d'analyse réconciliant la tradition du χ^2 issue de Karl Pearson et celle des *odds ratios* et des modèles log-linéaires, cet article majeur met en perspective une série de travaux de l'auteur portant sur les modèles d'association et l'analyse des correspondances, et propose un indice qui est, dans la tradition de l'*odds ratio*, l'équivalent du coefficient de contingence carrée moyenne φ^2 ou du V de Cramér dans la tradition pearsonienne.

Cet indice, noté λ par l'auteur, peut être dénommé coefficient d'association intrinsèque⁸, pour des raisons qui, en même temps que le détail de sa définition et les démonstrations associées, sont développées en annexe B (p. 525). Notons simplement ici qu'il est défini de manière relativement simple, comme l'écart-type des coefficients d'interaction log-linéaires

7 Ils rejoignent en cela une longue tradition d'indices se concentrant sur la diagonale de la table, comprenant par exemple l'indice d'*intrinsic status inheritance* (Goodman, 1969).

8 Dans la littérature portant sur les modèles log-linéaires, le qualificatif d'« intrinsèque » dénote une association indépendante des marges de la table, correspondant donc au concept d'homogamie relative. Cet indice ne doit pas être confondu avec l'indice λ de Goodman et Kruskal, avec lequel il n'entretient aucun lien.

entre lignes et colonnes (notés plus haut λ_{hf}^{HF} dans le modèle de stabilité de l'association). Il offre ainsi une mesure de l'écart à la situation d'indépendance, de l'ampleur des contrastes entre les sur- et les sous-représentations des différentes configurations de couple possibles, ou encore de l'écart des *odds ratios* à leur valeur de référence 1. Contrairement aux indices présentés plus haut, il ne privilégie donc en aucune manière la diagonale de la table. Détaché de son cadre d'origine lié aux modèles d'association, le coefficient d'association intrinsèque peut être utilisé comme indice général de l'association observée dans une table de contingence, en dehors de toute modélisation ; mais aussi éventuellement en combinaison avec d'autres familles de modèles, comme nous le verrons plus bas.

L'apport de cet article de Leo Goodman (1996) nous paraît ainsi de première importance, et l'on ne peut que regretter que l'auteur n'ait pas plus explicitement proposé le coefficient d'association intrinsèque comme mesure générale de l'association indépendante des marges de la table. En fournissant un standard de mesure de l'association, cet indice nous semble pouvoir faciliter la comparaison des travaux, très divers, qui recourent d'une manière ou d'une autre aux *odds ratios*. Signe de l'intérêt pratique du coefficient d'association intrinsèque, l'élaboration hautement théorique proposée par Goodman a été déjà mobilisée dans des analyses empiriques de premier plan, grâce à une extension directe de l'indice à la régression logistique (Hout, Brooks et Manza, 1995 ; principe repris par Breen *et al.*, 2009) . Néanmoins, cet usage semble être resté assez confidentiel, et ne s'est en particulier pas développé dans le domaine de l'analyse log-linéaire et log-multiplicative, qui était pourtant son univers d'origine.

2.3 L'*odds ratio* absolu moyen (MAOR)

Si le coefficient d'association intrinsèque nous paraît ainsi un outil particulièrement utile pour l'analyse de l'homogamie relative, il présente un inconvénient non négligeable dans l'optique de la communication des résultats des analyses log-linéaires : malgré ses qualités mathématiques, cet indice demeure relativement obscur par sa signification – l'écart-type de coefficients d'interaction log-linéaires –, son échelle – le logarithme des *odds ratios* – et son appellation même. Cette complexité constitue un obstacle même pour les spécialistes des méthodes fondées sur l'*odds ratio*.

C'est pourquoi nous voudrions ici proposer, en nous fondant sur le cadre théorique établi par Goodman, un indice apparenté au coefficient d'association intrinsèque, mais défini exclusivement à partir des *odds ratios* de la table, et exprimé dans leur échelle. En effet, l'*odds ratio* offre une interprétation en termes de chances relatives assez directe à partir des proportions observées dans les cellules de la table, et s'est imposé dans la communauté scientifique à travers notamment la régression logistique. Une association exprimée en *odds ratios* est donc plus aisément compréhensible, et son ordre de grandeur assimilable, que son équivalent mesuré sur une échelle logarithmique.

Ce nouvel indice, dénommé *odds ratio* absolu moyen (MAOR, pour *Mean Absolute Odds Ratio*), est défini comme l'écart-type (géométrique) de tous les *odds ratios* pouvant être construits à partir de la table. Il est égal à l'*odds ratio* dans le cas particulier d'une table à deux lignes et deux colonnes, et constitue une généralisation de cette quantité à des tables à deux dimensions de taille arbitraire. Comme le coefficient d'association intrinsèque, auquel il est lié de manière très directe, le MAOR mesure le contraste général entre les sur- et sous-représentations des différentes configurations de couples de la table. Indépendant des marges de la table comme l'*odds ratio*, il peut comme le χ^2 être décomposé en contributions par cellule, par ligne et par colonne. Finalement, si le MAOR est utilisable de manière non paramétrique sur une table d'effectifs observés, il est néanmoins lié de manière très directe à l'association spécifiée par différentes familles de modèles, comme le modèle UNIDIFF et les modèles d'association. Nous renvoyons encore une fois le lecteur à l'annexe B (p. 525) pour la présentation du cadre d'analyse élaboré par Goodman, ainsi que les démonstrations relatives au calcul du MAOR, à son lien avec le coefficient d'association intrinsèque et avec les modèles log-multiplicatifs cités.

Nous utiliserons par la suite la version pondérée du MAOR, en prenant comme poids les proportions marginales moyennes des différentes catégories, c'est-à-dire leur poids moyen dans la population sur la période étudiée. Cette pondération présente l'intérêt d'attribuer une importance variable aux groupes sociaux selon leur taille, impliquant ainsi qu'une évolution de l'homogamie concernant des groupes très nettement minoritaires ne peut pas compenser des mouvements inverses affectant la majorité de la société. Il s'agit ainsi de limiter l'influence du choix d'une classification particulière : sans le recours à une pondération par la moyenne des marges, diviser un groupe en deux sous-catégories

doublerait son poids dans l'intensité générale de l'association. En retenant comme poids les proportions moyennes sur la période, plutôt que les proportions observées chaque année, nous préservons en outre la propriété d'indépendance des marges pour l'étude des évolutions temporelles. Au total, la pondération par la moyenne des marges constitue un moyen terme qui nous paraît optimal entre l'absence de pondération (ou le recours à une pondération uniforme) et la pondération marginale année par année, qui détruit l'indépendance des marges⁹.

3 Un affaiblissement de l'homogamie relative

Les fondements méthodologiques posés, cette section vise à mesurer l'évolution de l'homogamie relative de diplôme, de classe et d'origine sociales en France au cours des quarante dernières années. Nous commençons par mettre en évidence l'affaiblissement d'ensemble de l'homogamie relative à l'aide du modèle UNIDIFF, avant de détailler cette évolution en étudiant séparément l'endogamie relative des différents groupes.

3.1 Une diminution de l'intensité générale de l'homogamie relative

Comme on pouvait s'y attendre, le modèle d'indépendance conditionnelle – qui fait l'hypothèse qu'il n'existe aucune association entre les caractéristiques des conjoints au-delà de ce qui serait observé si le choix du conjoint était aléatoire – est très clairement rejeté par nos données pour les trois composantes de l'homogamie que nous étudions, ainsi que l'indiquent les valeurs positives très élevées des indicateurs BIC et AIC. Cependant, seuls un peu plus d'un quart des couples ne sont pas classés correctement par ce modèle en ce qui concerne le diplôme et la classe sociale, et 17 % en ce qui concerne l'origine sociale (indice de dissimilarité Δ dans le tableau 2.1) : ce constat met en évidence l'influence majeure – et assez logique – de la structure de la population sur le choix du conjoint.

⁹ Ces deux options sont retenues par Breen *et al.* (2009, p. 1511), la première pour la mesure des évolutions indépendamment des marges, la seconde (sous une forme légèrement différente) pour indiquer le niveau général d'inégalité à chaque point temporel.

Tableau 2.1 : Indicateurs d'ajustement des modèles concernant l'ensemble des configurations de couple

	D. L.	Déviante	Δ (%)	BIC	AIC
Diplôme :					
Indépendance cond.	4 200	216 761	26,14	162 175	208 361
Association stable	4 100	10 336	5,04	-42 950	2 136
UNIDIFF	4 059	8 780	4,47	-43 974	662
UNIDIFF, linéaire	4 099	8 901	4,51	-44 373	703
Forme de régression	3 960	5 425	3,44	-46 042	-2 495
Forme de régression, linéaire	4 000	5 611	3,59	-46 377	-2 389
Classe sociale :					
Indépendance cond.	4 200	233 526	26,93	179 359	225 126
Association stable	4 100	7 594	4,15	-45 283	-606
UNIDIFF	4 059	6 683	3,77	-45 665	-1 435
UNIDIFF, linéaire	4 099	7 072	3,97	-45 792	-1 126
Forme de régression	3 960	5 588	3,37	-45 483	-2 332
Forme de régression, linéaire	4 000	5 720	3,43	-45 867	-2 280
Classe sociale des actifs :					
Indépendance cond.	4 200	184 919	29,75	132 560	176 519
Association stable	4 100	7 635	4,93	-43 478	-565
UNIDIFF	4 059	6 009	4,20	-44 593	-2 109
UNIDIFF, linéaire	4 099	6 178	4,30	-44 922	-2 020
Forme de régression	3 960	5 149	3,79	-44 218	-2 771
Forme de régression, linéaire	4 000	5 269	3,86	-44 598	-2 731
Classe sociale d'origine :					
Indépendance cond.	2 430	65 249	17,12	34 661	60 389
Association stable	2 349	3 582	3,96	-25 986	-1 116
UNIDIFF	2 320	2 992	3,46	-26 211	-1 648
UNIDIFF, linéaire	2 348	3 057	3,53	-26 499	-1 639
Forme de régression	2 240	2 691	3,22	-25 505	-1 789
Forme de régression, linéaire	2 268	2 774	3,33	-25 774	-1 762

D. L. : degrés de liberté. Δ : indice de dissimilarité.

Le modèle de stabilité, qui décrit parfaitement l'association entre les caractéristiques des conjoints, mais suppose qu'elle est restée constante au cours du temps, constitue une amélioration radicale du modèle précédent : supposer la stabilité de l'association permet de répartir correctement autour de 95 % des effectifs de nos tables d'homogamie. Gardons cependant à l'esprit que, même sans décrire du tout l'homogamie, le modèle d'indépendance conditionnelle classait déjà correctement la grande majorité des couples : même si l'homogamie relative avait complètement disparu sur la période, le modèle de stabilité resterait de bonne qualité selon cet indicateur.

Nous utilisons donc le modèle UNIDIFF pour mesurer l'évolution du niveau de l'homogamie en supposant que sa structure est restée la même (*cf.* l'encadré 2.1, p. 136, pour l'équation du modèle). La nette diminution des valeurs du BIC et de l'AIC indique que ce modèle doit de loin être préféré à celui de stabilité pour toutes les composantes¹⁰. Les *odds ratios* absolus moyens mesurant l'intensité globale de l'association chaque année sont présentés dans la figure 2.1.

Toutes les composantes ont connu un net affaiblissement de l'association, avec des ampleurs similaires mais des rythmes légèrement différents. L'homogamie de diplôme et celle de classe sociale sont de niveaux très proches sur toute la période, alors que celle d'origine sociale est nettement plus faible (phénomène en partie dû à l'information moins précise dont nous disposons sur cette dernière). Si, comme attendu, la prédominance des caractéristiques acquises sur les caractéristiques héritées est bien visible, en revanche, celle des facteurs culturels sur les facteurs socioprofessionnels est clairement rejetée sur la période étudiée – contredisant l'hypothèse de M. Forsé et L. Chauvel (1995) évoquée au chapitre précédent.

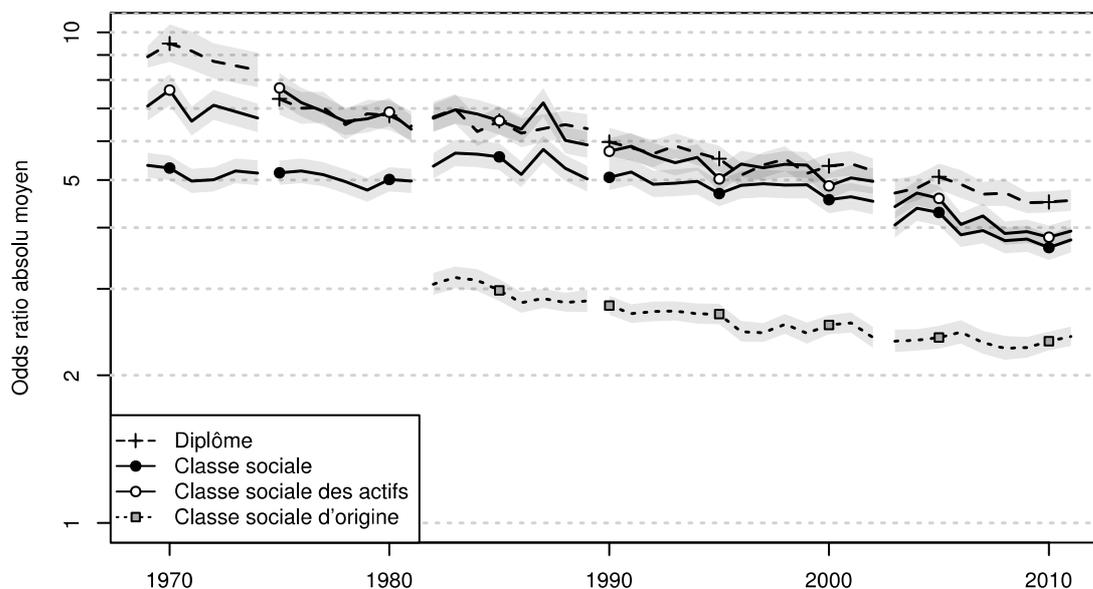
L'homogamie de diplôme, tout d'abord, a décliné de manière continue depuis 1969, avec une régularité remarquable. La baisse totale est de l'ordre de 40 à 50 % de l'association sur l'échelle des *odds ratios*, même après avoir soustrait la rupture de série de 1975.

L'homogamie de classe sociale, de son côté, semble être restée relativement stable entre 1969 et le milieu des années 1980, date à laquelle elle a commencé à décroître à un rythme soutenu qui se maintient jusqu'à aujourd'hui. Cette baisse semble légèrement plus rapide

10 Des modèles UNIDIFF plus contraints supposant une tendance temporelle linéaire sont à retenir selon le BIC, mais pas selon l'AIC, signe que la linéarité des évolutions est seulement partielle.

que celle de l'homogamie de diplôme, si bien que dans la période récente l'homogamie de classe sociale est légèrement plus faible que la première – phénomène qui pourrait s'accroître à l'avenir, validant avec retard l'hypothèse de M. Forsé et L. Chauvel discutée ci-dessus. La temporalité de cette évolution coïncide de manière frappante avec celle généralement retenue concernant le déclin du sentiment d'appartenance à la classe ouvrière (Michelat et Simon, 2004) aussi bien que du discours portant sur les classes sociales et les catégories socioprofessionnelles (Chauvel, 2006 ; Pierru et Spire, 2008).

Figure 2.1 : Évolution de l'intensité de l'homogamie d'après le modèle UNIDIFF



Source : enquêtes Emploi (Insee), 1969-2011

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans (bi-actifs dans le cas de la classe sociale des actifs)

Un *odds ratio* absolu moyen (cf. p. 143) de 1 correspond à la situation de choix aléatoire du conjoint (aucune homogamie). L'échelle verticale logarithmique permet la comparaison des écarts entre valeurs (équivalents à des rapports entre *odds ratios*). Les bandes indiquent des quasi-intervalles de confiance normaux à 95 % obtenus par bootstrap.

La série correspondant à la classe sociale en incluant les inactifs présente une rupture en 1982, qui peut être attribuée au meilleur codage de la dernière profession des inactifs suivant l'introduction de la nouvelle classification des PCS. Aussi bien avant qu'après cette rupture, l'homogamie de classe sociale apparaît légèrement plus forte pour les couples composés de deux actifs, ce qui est cohérent avec l'idée selon laquelle la dernière profession n'est pas un aussi bon indicateur de la position sociale des inactives que l'est la profession actuelle des actives. L'écart entre les deux séries se réduit avec le temps, une évolution attendue du fait de la diminution de la proportion d'inactives.

L'homogamie de classe sociale d'origine suit une tendance décroissante similaire depuis 1982, même si elle pourrait s'être stabilisée dans la période récente. Puisque les données ne sont pas disponibles avant 1982, nous ne pouvons pas observer si la baisse a commencé plus tôt, à l'instar de l'homogamie de diplôme, ou si elle n'a commencé que dans les années 1980, à l'instar de l'homogamie de classe sociale.

Dans l'ensemble, malgré une certaine inertie, les trois composantes essentielles de l'homogamie ont incontestablement perdu de leur importance – de l'ordre d'un tiers à la moitié de l'association en termes d'*odds ratios*.

3.2 Une baisse de l'endogamie relative de la plupart des groupes

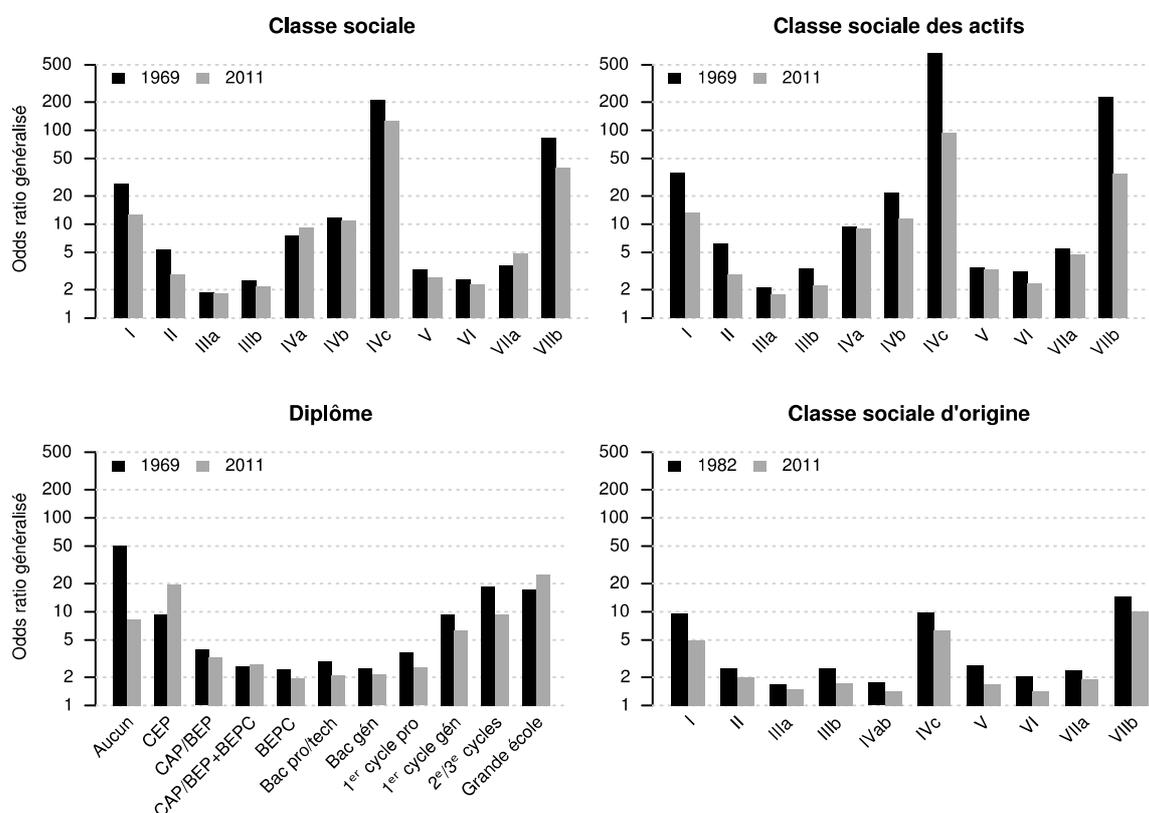
La tendance globale identifiée par les modèles UNIDIFF ne prend pas en compte les transformations de la structure de l'association entre les caractéristiques des conjoints. Comme on peut le voir dans le tableau 2.1, nos données sont assez riches pour aller au-delà de cette lecture, et décrire les changements concernant des configurations de couple particulières à l'aide du modèle à forme de régression (*cf.* l'encadré 2.1, p. 136, pour l'équation du modèle) : même si ce modèle n'est pas meilleur qu'UNIDIFF selon le BIC (à part dans le cas du diplôme), il l'est selon l'AIC¹¹. La légère diminution des indices de dissimilarité permet de conclure que la structure de l'homogamie n'a pas changé de manière radicale sur la période, mais que certaines transformations peuvent néanmoins être observées.

11 Dans tous les cas, supposer que la tendance est linéaire est justifié selon le BIC, mais pas selon l'AIC, ce qui peut encore une fois être considéré comme le signe que la tendance est essentiellement, mais pas complètement, linéaire (coefficients non repris ici).

Nous recourons ici encore à l'*odds ratio* généralisé (Kaufman et Schervish, 1987) pour présenter l'intensité de l'endogamie relative des différents groupes aux états de départ et d'arrivée. Comme au chapitre précédent, ceux-ci mesurent les chances d'un individu de ce groupe de se mettre en couple à l'intérieur du groupe plutôt qu'en dehors, par rapport aux chances équivalentes d'un individu non membre de ce groupe. La différence tient ici au fait que ces *odds ratios* sont calculés non plus directement à partir des données observées, mais à partir du modèle à forme de régression (*cf.* l'encadré 2.1, p. 136, pour le détail de leur calcul). Le recours à ce modèle permet une plus grande précision, puisque les données relatives à l'ensemble des enquêtes sont mobilisées : l'état de départ (respectivement, l'état d'arrivée) ne correspond pas strictement à l'année 1969 ou 1982 (respectivement, 2011), mais à ce que l'évolution observée sur l'ensemble de la période conduit à inférer pour ces années. Cette plus grande précision s'obtient au prix d'une hypothèse relativement faible, selon laquelle les autres années sont correctement décrites comme un état intermédiaire entre celui de départ et celui d'arrivée.

Ces *odds ratios* généralisés, mesurant l'endogamie aux états de départ et d'arrivée, sont présentés dans la figure 2.2. Un résultat vérifié pour l'ensemble de la période et toutes les composantes de l'homogamie est le fait bien connu que les groupes situés aux deux extrêmes de l'échelle sociale sont plus endogames (en termes relatifs) que ceux situés au milieu (structure en forme de U, voir Kalmijn, 1994, p. 444). Ce constat doit être nuancé en ce qui concerne la classe sociale, en mettant de côté les indépendants, dont l'endogamie est notoirement très forte. Un autre point notable tient à ce que, mis à part les ouvriers agricoles (VIIb), les catégories supérieures sont nettement plus endogames que celles situées en bas de l'échelle.

Figure 2.2 : Odds ratios généralisés mesurant l'endogamie dans les modèles à forme de régression



Source : enquêtes Emploi (Insee), 1969-2011

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans (et les deux conjoints ont déjà travaillé dans le cas de la classe sociale)

Un *odds ratio* généralisé compare les chances de former un couple endogame à celles de former n'importe quel autre type de couple. Une valeur de 1 correspond à une endogamie relative nulle. L'échelle verticale logarithmique permet la comparaison des écarts entre valeurs (équivalents à des rapports entre *odds ratios*).

En ce qui concerne l'endogamie de diplôme, le résultat le plus frappant tient à ce que la baisse concerne presque toutes les catégories. Les deux plus fortes diminutions concernent les non diplômés et les titulaires d'un diplôme universitaire des 2^e et 3^e cycles¹², ce qui n'empêche pas ces groupes de demeurer parmi les plus endogames. Les deux seules augmentations concernent les titulaires du Certificat d'études et les diplômés des grandes écoles. Le premier cas est dû à la disparition progressive du CEP, qui est devenu un diplôme passé exceptionnellement et uniquement à l'âge adulte avant de disparaître.

12 Toutes les évolutions commentées ici sont statistiquement significatives au seuil de 5 %.

Le second cas est moins banal. L'endogamie des diplômés des grandes écoles dépasse désormais celle du groupe des cadres et professions intellectuelles supérieures. En étudiant dans le détail les *odds ratios* du modèle à forme de régression (non repris ici), nous avons pu observer que les hommes comme les femmes diplômés des grandes écoles ont eu de plus en plus fortement tendance à choisir un conjoint dans leur propre groupe, mais aussi parmi les titulaires d'un diplôme professionnel du 1^{er} cycle ou d'un diplôme universitaire des 2^e et 3^e cycles, et de moins en moins à choisir un conjoint dans la plupart des autres groupes¹³. Nous savons que, contrairement aux universités, les grandes écoles n'ont pas connu de démocratisation de leur recrutement : la part des enfants issus des classes populaires parmi les élèves de ces écoles a baissé plus rapidement que dans la population générale (Euriat et Thélot, 1995 ; Albouy et Wanecq, 2003 ; Ichou et Vallet, 2013 à propos de l'opposition entre universités et classes préparatoires). L'évolution de l'endogamie en termes de diplôme que nous observons pour ce groupe reflète donc probablement, au moins en partie, son endogamie du point de vue de l'origine sociale. Un second facteur a pu jouer : alors que les grandes écoles étaient réservées aux hommes pour les cohortes les plus anciennes, les femmes les ont depuis intégrées de plus en plus fréquemment, quoique dans une mesure limitée (cf. chapitre précédent) : cette évolution a rendu possible la transformation de ces écoles en marchés conjugaux¹⁴. Les grandes écoles sont aujourd'hui des institutions particulièrement propices aux rencontres endogames du fait de l'intense sociabilité qui se développe sur leurs campus (Masse, 2002), qui peuvent être considérés selon la tripartition des lieux de rencontre élaborée par Michel Bozon et François Héran (1988, p. 125) comme des lieux réservés.

L'endogamie de classe sociale a décliné pour tous les groupes si nous considérons uniquement les actifs. Les deux seules exceptions concernent les indépendants avec salariés (IVa) et les techniciens (V), dont l'endogamie est restée stable. Les plus fortes baisses concernent le monde agricole (agriculteurs, IVc ; ouvriers agricoles, VIIb), ainsi que les

13 Nous avons vérifié que cette évolution peut être observée séparément dans chaque série d'enquête : elle n'est donc pas due à un changement dans le codage de cette catégorie. Notons qu'Henri Mendras (1988, p. 52) avait souligné la constitution d'« isolats professionnels » parmi les plus diplômés ; il ne mentionnait cependant pas les grandes écoles.

14 Cette évolution a d'ailleurs été à l'origine d'attaques à l'égard des femmes entrées à Sciences Po, que l'on a longtemps soupçonnées de fréquenter l'institution dans le seul objectif de rencontrer un bon parti. Cette école remplit aujourd'hui objectivement ce rôle pour les étudiants des deux sexes : le magazine *L'Imparfait* (n°1, 2009) estimait à partir d'un échantillon non représentatif de 300 étudiants à 25 % le taux d'endogamie à Sciences Po Paris – et même 53 % pour les enfants de diplômés du même établissement.

cadres et professions intellectuelles (I et II). L'ouverture conjugale des agriculteurs (Giraud et Rémy, 2008) reflète l'extension de leur « *aire de recrutement du conjoint* » et leur poids moins important dans les lieux de rencontre classiques comme le bal à partir des années 1960 (Bozon et Héran, 1987a, 1987b, p. 953-955). On peut penser que celle des cadres est due à la croissance des effectifs du groupe et à l'augmentation de l'hétérogénéité des origines sociales qui en a découlé (Vallet, 2011), ainsi que plus largement à l'augmentation de la diversité interne et à un certain brouillage des frontières de cette catégorie (Bouffartigue, 1999 ; Bouffartigue, Gadea et Pochic, 2011).

Si nous considérons les inactifs aussi bien que les actifs, on observe une évolution similaire, à l'exception des indépendants avec salariés (IVa) et des ouvriers non qualifiés (VIIa), dont l'endogamie a augmenté. Concernant les indépendants, la différence entre actifs et inactifs pourrait être due à un effet de déclaration, si le nombre de salariés de l'entreprise est déclaré plus précisément qu'auparavant pour les aides familiaux¹⁵. Concernant les ouvriers non qualifiés, elle peut être expliquée par une évolution du sens de cette appartenance pour les femmes : au début de la période, les inactives considérées comme ouvrières se sont en fait retirées du marché du travail tôt dans leur carrière ; au contraire, nous pouvons faire l'hypothèse que dans les années récentes, le fait d'être ouvrière non qualifiée, correspondant plus souvent à une profession exercée tout au long de la carrière, est un marqueur bien plus fort du statut social.

Enfin, l'endogamie selon la classe sociale d'origine est toujours plus faible que celle suivant la classe sociale des conjoints, et elle a décliné pour l'ensemble des groupes sans exception. La baisse la plus importante concerne les cadres et professions intellectuelles supérieures (I). On peut l'interpréter comme la conséquence de l'augmentation progressive au fil des cohortes de la fréquence de la mobilité descendante parmi les enfants de cadres, et notamment ceux de « cadres de promotion » aux origines populaires (Peugny, 2008, 2011). Néanmoins, ce groupe demeure parmi les plus endogames, venant juste après les ouvriers agricoles (VIIb) et les agriculteurs (IVc).

15 Cf. chapitre précédent et annexe A (p. 517).

4 Les transformations de l'espace social

Les modèles RC et RC-L portant sur les configurations exogames (diagonale de la table exclue) offrent un complément à cette description de l'endogamie, en synthétisant en plusieurs dimensions la structure sociale et ses évolutions au cours du temps. Nous commençons par introduire ces modèles et leur interprétation, avant de présenter les résultats.

4.1 Les modèles RC et RC-L

Les modèles log-multiplicatifs d'association lignes-colonnes, dits RC (pour *row-column association model*, dits encore modèles de type II de Goodman), sont l'équivalent de l'analyse factorielle des correspondances (AFC) dans la tradition de l'*odds ratio* et de la modélisation log-linéaire (Goodman, 1979, 1985, 1986, 1991 ; Wong, 2010). Les modèles RC-L (pour *row-column with layer effect* ; Clogg, 1982 ; Becker et Clogg, 1989 ; Wong, 2001, 2010) sont une généralisation de ces modèles à l'analyse de tables de contingence à trois dimensions. Moins restrictifs que le modèle UNIDIFF, qui suppose que la structure de l'homogamie est restée la même au cours du temps, ces modèles sont généralement¹⁶ plus parcimonieux que le modèle à forme de régression. Néanmoins, ils ne décrivent pas nécessairement moins bien les données : en effet, ils autorisent les différentes dimensions de l'espace social à évoluer à des rythmes différents, alors que le modèle à forme de régression suppose un rythme commun à toutes les configurations de couple (*cf.* l'encadré 2.1, p. 136 pour l'équation de ces modèles).

Supposant une certaine familiarité du lecteur avec l'analyse des correspondances, nous présentons ici les principales particularités des modèles RC et RC-L par rapport à cette technique. Le lien entre ces derniers et l'analyse des correspondances a été longuement développé par plusieurs auteurs (par exemple, Goodman, 1986, 1996 ; van der Heijden, Mooijaart et Takane, 1994). Le point commun de ces deux méthodes est de synthétiser en un certain nombre de dimensions, ordonnées en fonction de leur importance, l'association existant entre les modalités de deux variables catégorielles, définie comme écart à l'hypothèse d'indépendance. Pour cela, des scores ou coordonnées sont attribués, sur chaque

16 Il est toujours possible, en introduisant un nombre très élevé de dimensions, d'obtenir des modèles RC-L d'une très grande complexité (*cf.* l'encadré 2.1, p. 136) ; cependant, il est très rare de retenir plus de trois dimensions dans un tel modèle.

dimension ou axe, à chacune des modalités lignes et colonnes ; à chaque dimension est attribué un indice mesurant l'ampleur de l'association qu'elle décrit.

Ainsi, les dimensions identifiées par les modèles RC et RC-L s'interprètent de la même manière que les axes d'une analyse des correspondances, à l'exception du fait que, dans le cas des modèles RC-L, l'importance des dimensions (ici mesurée par l'*odds ratio* absolu moyen¹⁷) varie pour chaque année d'enquête. Se trouvent du même côté d'un axe les catégories hommes et les catégories femmes qui sont plus fréquemment associées, au sein des couples dans lesquels les conjoints appartiennent à des catégories différentes, que ce ne serait le cas en situation de choix aléatoire du conjoint ; à l'inverse, les catégories négativement associées se trouvent sur des côtés opposés des axes. Plus une catégorie est éloignée de l'origine des axes, et plus elle présente un choix du conjoint contrasté, c'est-à-dire inégalement réparti entre les catégories du sexe opposé. Deux catégories très éloignées de l'origine sont fortement associées si elles sont du même côté d'un axe, et très négativement associées si elles sont de côtés différents.

Une première différence, essentielle ici pour l'étude de l'homogamie relative, tient au fait que les modèles RC et RC-L, fondés sur l'*odds ratio*, sont indépendants des transformations marginales de la table lorsque des pondérations elles-mêmes indépendantes de ces marges sont utilisées – au contraire de l'analyse des correspondances, qui est fondée sur la métrique du χ^2 .

Une seconde différence tient au choix du nombre de dimensions à analyser. Un modèle RC ou RC-L comprend un nombre de dimensions déterminé à l'avance ; on retient le modèle qui s'ajuste le mieux aux données en suivant divers critères d'inférence statistique, qui sont habituellement le test du rapport de vraisemblance, le BIC ou l'AIC. Le nombre de dimensions à retenir dépend de ce choix ; l'importance de ces dimensions est donnée par le coefficient d'association intrinsèque de chaque dimension, ou de manière équivalente par son *odds ratio* absolu moyen associé. Notons cependant que contrairement à la part de variance expliquée utilisée en analyse des correspondances, ces deux valeurs donnent une indication concernant l'intensité absolue de l'association, c'est-à-dire qu'elles peuvent être directement comparées à celles obtenues à partir d'autres tables.

17 Cf. l'annexe B (p. 525) pour la démonstration du lien entre le MAOR et les modèles RC-L.

Finalement, les modèles RC-L permettent de modéliser de manière parcimonieuse l'évolution temporelle en attribuant pour chaque année un coefficient d'association intrinsèque différent à chaque dimension. Ils permettent donc de résumer l'espace social en supposant que celui-ci peut être décrit par un certain nombre de dimensions sur lesquelles les groupes sociaux s'ordonnent ; la position (scores ou coordonnées) des catégories est supposée stable, mais l'importance des dimensions varie, certaines dimensions pouvant même apparaître, disparaître ou s'inverser (coefficient d'association intrinsèque négatif, ou *odds ratio* absolu moyen inférieur à 1).

Tableau 2.2 : Indicateurs d'ajustement des modèles concernant les configurations exogames

	D. L.	Déviante	Δ (%)	BIC	AIC
Diplôme :					
Indépendance cond.	3 738	95 445	20,97	48 408	87 969
Association stable	3 649	6 766	5,16	-39 151	-532
UNIDIFF	3 608	6 372	4,95	-39 029	-844
UNIDIFF, linéaire	3 648	6 428	4,97	-39 476	-868
RC(1)-L	3 678	9 922	6,22	-36 360	2 566
RC(2)-L	3 618	7 072	5,18	-38 455	-164
RC(3)-L	3 558	5 508	4,33	-39 264	-1 608
RC(1)-L linéaire	3 718	9 985	6,24	-36 800	2 549
RC(2)-L linéaire	3 698	7 226	5,30	-39 308	-170
RC(3)-L linéaire	3 678	5 768	4,51	-40 514	-1 588
Classe sociale :					
Indépendance cond.	3 738	81 124	18,69	33 940	73 648
Association stable	3 649	6 598	4,51	-39 462	-700
UNIDIFF	3 608	5 750	4,03	-39 793	-1 466
UNIDIFF, linéaire	3 648	5 950	4,17	-40 098	-1 346
RC(1)-L	3 678	24 902	7,06	-21 525	17 546
RC(2)-L	3 618	7 358	4,29	-38 311	122
RC(3)-L	3 558	5 482	3,76	-39 430	-1 634
RC(1)-L linéaire	3 718	25 101	7,16	-21 831	17 665
RC(2)-L linéaire	3 698	7 685	4,51	-38 994	289
RC(3)-L linéaire	3 678	5 941	4,15	-40 486	-1 415

	D. L.	Déviante	Δ (%)	BIC	AIC
Classe sociale des actifs :					
Indépendance cond.	3 738	58 985	19,47	13 537	51 509
Association stable	3 649	6 182	5,27	-38 183	-1 116
UNIDIFF	3 608	5 174	4,59	-38 693	-2 042
UNIDIFF, linéaire	3 648	5 295	4,70	-39 058	-2 001
RC(1)-L	3 678	19 720	9,21	-24 998	12 364
RC(2)-L	3 618	6 307	4,74	-37 681	-929
RC(3)-L	3 558	5 052	4,48	-38 207	-2 064
RC(1)-L linéaire	3 718	19 922	9,30	-25 283	12 486
RC(2)-L linéaire	3 698	6 579	5,01	-38 383	-817
RC(3)-L linéaire	3 678	5 407	4,76	-39 312	-1 949
Classe sociale d'origine :					
Indépendance cond.	2 130	16 811	10,34	-9 348	12 551
Association stable	2 059	2 620	3,81	-22 666	-1 498
UNIDIFF	2 030	2 491	3,67	-22 439	-1 569
UNIDIFF, linéaire	2 058	2 550	3,74	-22 725	-1 566
RC(1)-L	2 084	5 081	5,81	-20 513	913
RC(2)-L	2 038	2 527	3,65	-22 502	-1 549
RC(3)-L	1 992	2 390	3,54	-22 074	-1 594
RC(1)-L linéaire	2 112	5 128	5,85	-20 809	904
RC(2)-L linéaire	2 094	2 623	3,77	-23 093	-1 565
RC(3)-L linéaire	2 076	2 524	3,70	-22 971	-1 628

D. L. : degrés de liberté. Δ : indice de dissimilarité.

Ainsi, l'intérêt de recourir à des modèles RC-L plutôt qu'au modèle UNIDIFF plus simple utilisé plus haut, ainsi que le nombre de dimensions à retenir, dépendent à la fois de la complexité de l'évolution de la structure sociale et de la richesse de nos données. À la lecture des indicateurs présentés dans le tableau 2.2, les modèles RC-L à deux ou trois dimensions, selon les cas, constituent une amélioration par rapport à UNIDIFF ; un nombre supérieur de dimensions ne paraît pas justifié¹⁸. Nous retenons donc un modèle à trois

18 Un modèle RC-L à trois dimensions – RC(3)-L – décrit mieux les configurations de couple exogames que le modèle UNIDIFF en ce qui concerne l'homogamie de diplôme, selon le BIC et l'AIC, et ce, que les tendances temporelles soient linéaires ou libres. En ce qui concerne l'homogamie de classe sociale, seul un

dimensions pour décrire l'homogamie de diplôme et de classe sociale, et un modèle à deux dimensions pour décrire l'homogamie d'origine sociale.

Si les modèles RC-L sont de puissants outils pour synthétiser les transformations à l'œuvre indépendamment de l'évolution de la composition de la population, il ne semble pas inutile de revenir dans un second temps à la structure de l'homogamie telle qu'on peut l'observer une année donnée, de manière à replacer ces transformations dans leur contexte et ainsi mieux percevoir leur signification. En outre, les modèles RC-L, qui se concentrent sur les mouvements d'ensemble, ne font pas toujours ressortir correctement des changements plus mineurs dans la position relative de quelques groupes sur une dimension, qui ont pourtant parfois un grand intérêt pour préciser les résultats. Ainsi, après avoir étudié les tendances à l'aide d'un modèle RC-L, nous présenterons pour chaque composante des instantanés de l'espace social à l'aide de modèles RC appliqués à l'homogamie observée en début et en fin de période. Nous utiliserons pour cela une pondération marginale qui supprime la propriété d'indépendance des marges de ces modèles.

Cette approche est similaire à celle d'Alain Desrosières (1978), qui recourait lui à l'analyse des correspondances. Néanmoins, les modèles d'association nous permettent ici d'exclure l'association endogame (diagonale de la table), déjà décrite plus haut, et d'obtenir ainsi un espace social plus lisible et reflétant uniquement les relations entre groupes¹⁹. Afin de saisir plus précisément ces associations, nous relient les points hommes et femmes par des lignes dont l'épaisseur et l'opacité reflètent la valeur de l'*odds ratio* généralisé correspondant à la configuration de couple correspondante²⁰. Ainsi se constitue un réseau faisant apparaître les différents blocs de l'espace social, et le cas échéant les passerelles faisant le lien entre eux. De manière à décrire l'ensemble des configurations, l'endogamie

modèle à trois dimensions postulant une tendance linéaire est meilleur selon le BIC, quoique l'AIC préfère lui des tendances non linéaires. Enfin, en ce qui concerne l'homogamie d'origine sociale, deux dimensions et des tendances linéaires sont suffisantes selon le BIC, mais l'AIC préfère des modèles plus complexes ; nous retenons la solution plus simple suggérée par le premier indicateur, quoique dans sa version non linéaire afin de permettre au lecteur de juger de la régularité des tendances.

19 On pourra se référer à nos deux articles méthodologiques (Bouchet-Valat, 2015b, 2015c) pour plus de détails concernant les limites de l'analyse des correspondances classique pour la représentation d'une table d'homogamie, et sa comparaison avec les modèles d'association.

20 Les associations négatives (répulsions) ne sont pas représentées : elles correspondent généralement à des points très éloignés sur le plan, rendant la figure très peu lisible ; en outre, elles ne sont que le reflet des associations positives (la moyenne géométrique des *odds ratios* étant par définition égale à 1 pour chaque catégorie).

relative est elle aussi représentée par des traits reliant les points homme et femme d'une même catégorie, même si elle n'intervient pas dans la construction du plan.

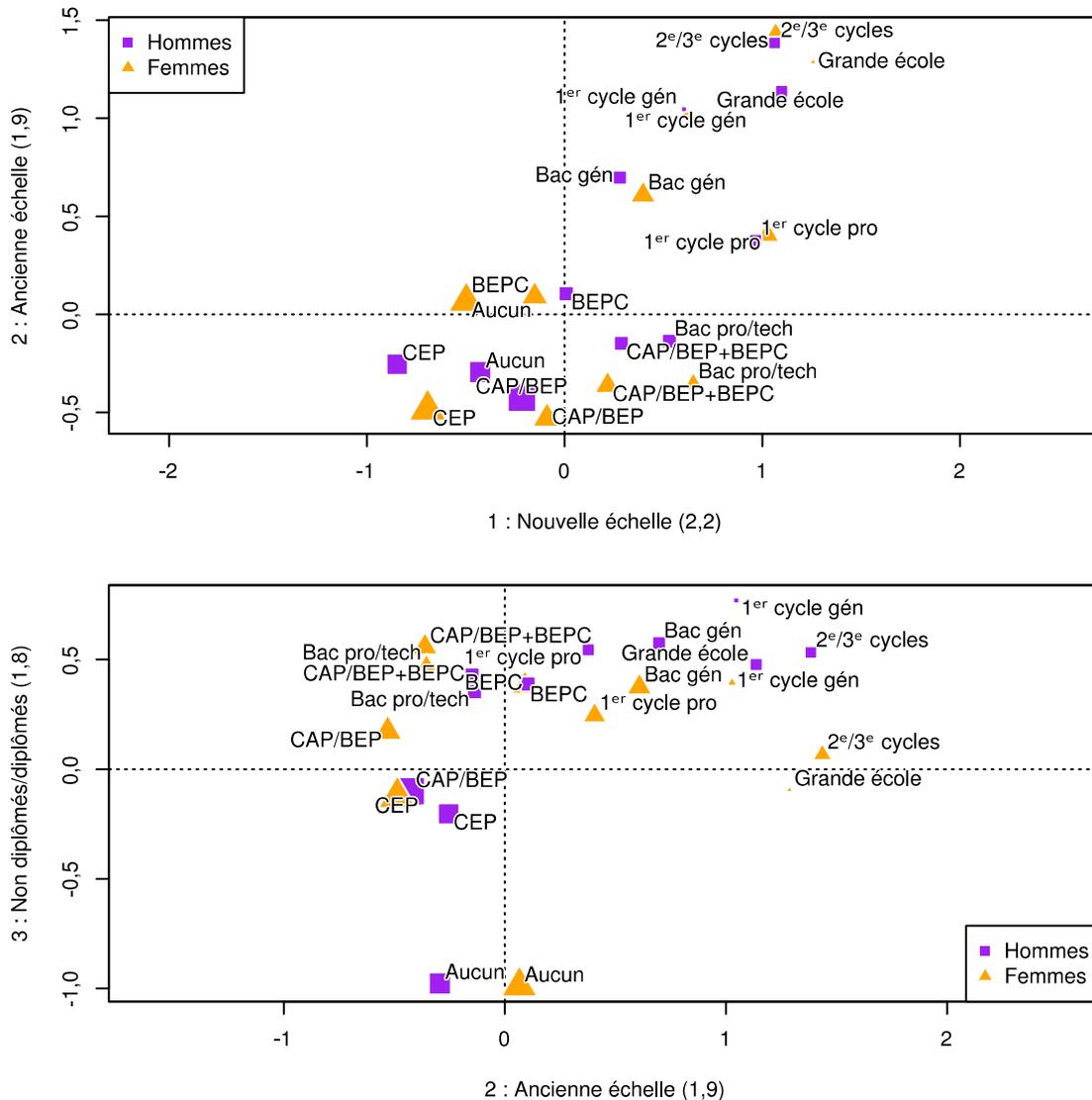
Nous retiendrons pour toutes les composantes un modèle RC à deux dimensions, ce qui permet une représentation sur un plan unique. Une fois exclue la diagonale de la table, des modèles à plus de deux dimensions se révèlent trop complexes pour les données, au point qu'ils ne peuvent être estimés correctement. De fait, la qualité d'ajustement des modèles à deux dimensions (reprise en légende des graphiques correspondants) s'avère très satisfaisante.

4.2 L'homogamie de diplôme : la fin de la spécificité des non diplômés

Le modèle RC-L décrivant le mieux l'homogamie de diplôme est selon le BIC celui comprenant trois dimensions. Décrivons tout d'abord ces dimensions (Figure 2.3) avant de résumer l'évolution globale qu'elles dépeignent, puis de représenter la structure de l'homogamie en début et en fin de période qui en découle. Un de leurs intérêts réside, particulièrement en ce qui concerne la troisième dimension, dans le fait qu'elles questionnent la conception courante selon laquelle l'éducation peut être correctement modélisée par une échelle unique, parfois mesurée par le nombre d'années d'études requises pour obtenir un diplôme donné – ce qui nie les différences de statut entre filières.

La première dimension ordonne de manière attendue les diplômes, du primaire jusqu'au supérieur. La seule exception à cette règle concerne les titulaires du Certificat d'études, qui se situent plus bas sur l'échelle que les non diplômés : ainsi que nous l'avons avancé plus haut, la perte de valeur conjugale de ce titre est probablement la conséquence de sa disparition progressive, qui l'a réservé à un public de plus en plus restreint (*cf.* chapitre précédent). Cette échelle reflète l'état vers lequel le système éducatif s'oriente dans les années récentes : en effet, cette dimension, qui était la plus faible des trois en 1969, a continûment gagné en importance, pour devenir la plus forte en 2011 (Figure 2.4).

Figure 2.3 : Dimensions du modèle RC(3)-L pour l'homogamie de diplôme



Source : enquêtes Emploi (Insee), 1969-2011

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Une catégorie homme et une catégorie femme sont d'autant plus éloignées de l'origine et du même côté d'un axe (respectivement d'un côté différent) que leur combinaison est sur-représentée (resp. sous-représentée) parmi les configurations exogames par rapport à la situation de choix aléatoire du conjoint.

L'intensité de chaque dimension (*odds ratio* absolu moyen entre parenthèses, cf. p. 143) correspond à la moyenne sur la période. Une pondération par la moyenne des marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes.

À l'inverse, la deuxième dimension a décliné au cours de la période, quoique lentement depuis les années 1980 : elle décrit donc la structure de l'homogamie qui prédominait dans les années 1960, même si elle n'a certainement pas disparu en 2011. Bien que très similaire à la première dimension, elle s'en distingue de manière non négligeable. Tout d'abord, les diplômes réservés aux classes populaires sont très proches les uns des autres, au bas de l'échelle, ce qui est le signe de choix du conjoint très semblables et d'une grande perméabilité entre ces catégories au début de la période. Ensuite, à chaque niveau, les diplômes professionnels sont situés plus bas que leurs équivalents de la filière générale sur cette dimension : à niveau égal, les diplômes professionnels étaient associés à un statut inférieur au début de la période ; au contraire, sur la première dimension, on n'observe soit aucune différence, soit une différence en faveur des diplômes professionnels, signe d'une tendance à la réduction de cet écart. Enfin, les hommes non diplômés sont situés nettement plus bas que les femmes de la même catégorie, ce qui est un signe du fait que ne posséder aucun diplôme n'était pas aussi marquant du point de vue conjugal pour les femmes que pour les hommes dans les années 1960. Plus généralement, les différences entre les positions des hommes et des femmes sont systématiquement plus importantes sur cette dimension que sur la première, ce qui indique de plus fortes asymétries de genre du point de vue du choix du conjoint en 1969 qu'en 2011.

Considérées ensemble, les deux premières dimensions dessinent aussi une opposition secondaire entre filière générale (dans la partie supérieure gauche du plan) et filière professionnelle (dans sa partie inférieure droite).

Finalement, la troisième dimension oppose les non diplômés à toutes les autres catégories de diplôme au-delà du BEP, sans considération de niveau ; les titulaires d'un Certificat d'études, d'un CAP ou d'un BEP constituent un groupe intermédiaire. Cette dimension était la plus importante en 1969, mais elle a régulièrement décliné depuis et a même presque complètement disparu en 2011. Les *odds ratios* du modèle à forme de régression (Figure 2.2) confirment cette évolution du point de vue de l'endogamie : la catégorie des non diplômés est celle pour laquelle la diminution a été la plus marquée. Ce résultat est important : on aurait pu s'attendre à ce que les non diplômés, constituant un groupe de plus en plus spécifique au cours du temps, fassent l'expérience de difficultés grandissantes pour trouver un conjoint, et plus généralement pour s'intégrer au reste de la

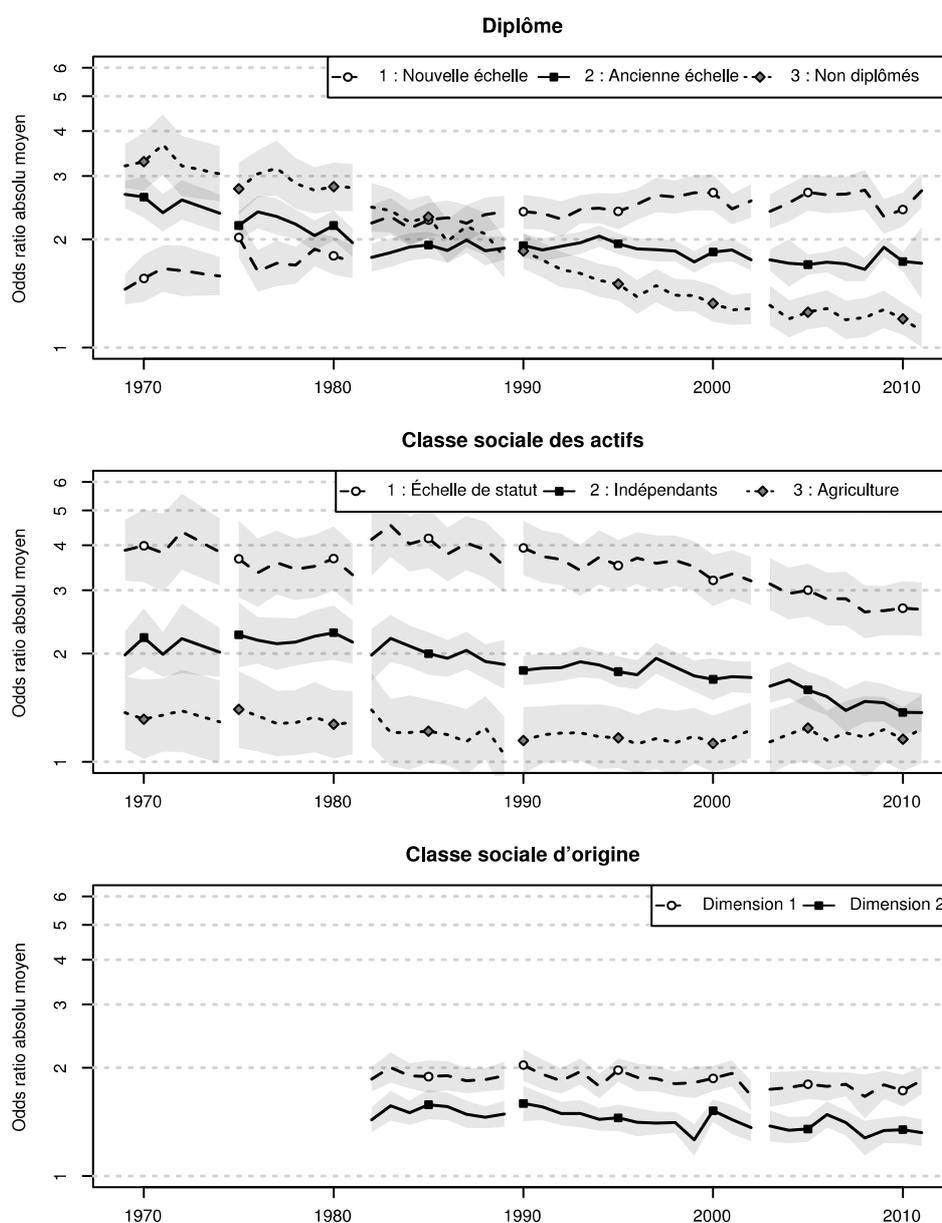
société²¹ ; ce n'est clairement pas ce que nous observons à travers l'homogamie. Au contraire, avec la massification scolaire, les non diplômés ont cessé de constituer un groupe social cohérent porteur d'une forte identité, pour devenir simplement le bas d'une hiérarchie. La disparition de la troisième dimension, qui était la seule à ne pas correspondre à un gradient de diplômes, mais à une opposition essentiellement binaire (diplômés/non-diplômés), semble marquer une évolution vers une stratification éducative faite de plus de continuités et de moins de ruptures – même si la distinction entre diplômés de l'enseignement général et diplômés de l'enseignement technique et professionnel, qui était déjà visible au début de la période, persiste.

Il est en outre important de souligner que l'échelle des diplômes n'a pas gagné en importance de manière à complètement compenser l'affaiblissement de l'opposition entre diplômés et non diplômés : tout au plus observe-t-on une recomposition de cette échelle (première et deuxième dimensions). Ce résultat est à mettre en relation avec les débats portant sur la démocratisation scolaire : en transposant la typologie proposée par Pierre Merle (2000), le scénario que nous observons semble être le reflet d'une démocratisation plus « égalisatrice » que « ségrégative », étant donné que les anciennes distinctions n'ont pas été remplacées par de nouvelles. Cette conclusion est très similaire à celle à laquelle aboutissent Louis-André Vallet et Marion Selz (2007) à propos de l'évolution des inégalités des chances devant l'école selon l'origine sociale.

Ce résultat mis en évidence, notons cependant que le déclin de l'homogamie de diplôme pourrait bien ralentir ou s'arrêter dans les années à venir, étant donné que l'opposition entre diplômés et non diplômés (troisième dimension), qui a fourni l'essentiel de la baisse (endogamie mise à part), a aujourd'hui complètement disparu (Figure 2.4).

21 C'est ce qu'observent Jeroen Smits et Hyunjoon Park (2009, p. 245) suite à la massification scolaire ayant pris place dans dix pays asiatiques.

Figure 2.4 : Évolution de l'intensité des dimensions des modèles RC-L

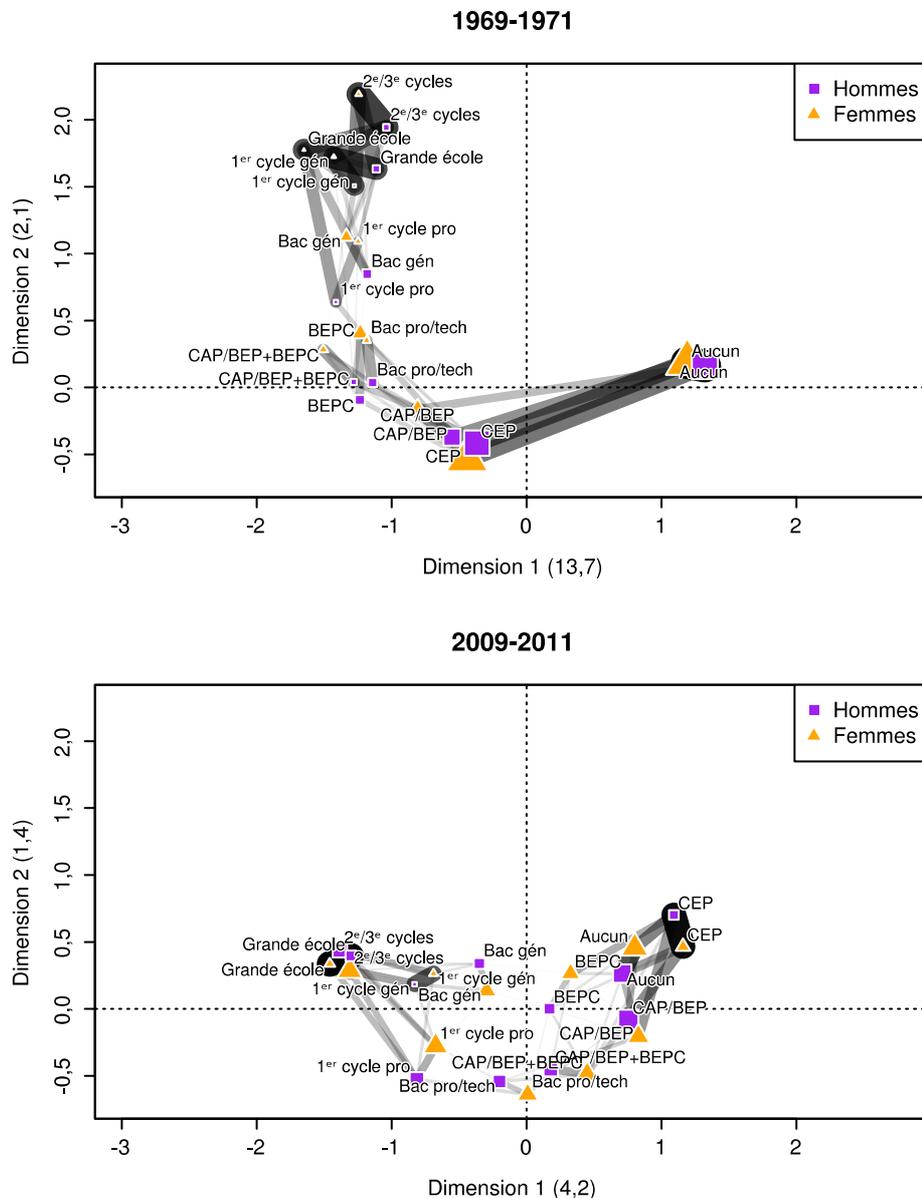


Source : enquêtes Emploi (Insee), 1969-2011

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Un *odds ratio* absolu moyen de 1 correspond à une dimension d'importance nulle (cf. p. 143). L'échelle verticale logarithmique permet la comparaison des écarts entre valeurs (équivalents à des rapports entre *odds ratios*). Les bandes indiquent des quasi-intervalles de confiance normaux à 95 % obtenus par bootstrap. Une pondération par la moyenne des marges est utilisée.

Figure 2.5 : L'homogamie de diplôme en 1969-1971 et en 2009-2011 selon le modèle RC(2)



Source : enquêtes Emploi (Insee)

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Une catégorie homme et une catégorie femme sont d'autant plus éloignées de l'origine et du même côté d'un axe (respectivement d'un côté différent) que leur combinaison est sur-représentée (resp. sous-représentée) parmi les configurations exogames par rapport à la situation de choix aléatoire du conjoint.

L'intensité de chaque dimension est indiquée par l'*odds ratio* absolu moyen entre parenthèses (cf. p. 143). Une pondération par les marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes. L'épaisseur et l'opacité des traits reliant les points hommes et femmes indiquent l'intensité de

l'association positive entre les catégories correspondantes (*odds ratio* généralisé); les associations négatives ne sont pas représentées.

Le modèle portant sur les années 1969-1971 repose sur 44 838 observations et présente une déviance de 216 pour 53 degrés de liberté, avec un indice de dissimilarité de 2,4 %. Celui portant sur les années 2009-2011 repose sur 41 584 observations et présente une déviance de 251 pour 53 degrés de liberté, avec un indice de dissimilarité de 3,0 %.

Quelle structure de l'espace social dessinent ces tendances en début et en fin de période ? Comme dans la figure 2.3, nous retrouvons en 1969-1971 (Figure 2.5) l'opposition très nette entre diplômés et non diplômés sur la première dimension (axe horizontal), alors que la seconde dimension (axe vertical) correspond à l'échelle des diplômes. Les liens tracés entre les points mettent en évidence deux ensembles bien distincts : l'un constitué autour des non diplômés ainsi que des titulaires du certificat d'études, du CAP ou du BEP ; l'autre autour des diplômés des grandes écoles et des filières générales de l'université. La rupture entre ces deux pôles s'effectue entre le baccalauréat professionnel (rattaché au secondaire inférieur et au primaire) et le baccalauréat général (rattaché au supérieur).

En 2009-2011, l'opposition entre diplômés et non diplômés a, comme on l'a déjà souligné, complètement disparu : c'est désormais l'échelle des diplômes qui constitue la première dimension (axe horizontal). La seconde dimension recouvre désormais une opposition entre diplômes professionnels (en bas de la figure) et diplômes généraux (en haut), à tous les niveaux de la hiérarchie. On peut supposer que cette distinction correspond autant à des origines et à des destinées sociales différentes qu'à l'effet sur les rencontres de l'organisation par filières du système scolaire. Son développement s'est accompagné d'une amélioration²² de la position, sur l'échelle des diplômes, des titres du premier cycle professionnel du supérieur et du baccalauréat professionnel, qui se trouvent désormais au même niveau que leurs équivalents du général. À l'inverse, signe de leur massification, certains diplômes généraux ont vu leur position baisser par rapport aux autres : les diplômes du premier cycle général se détachent des diplômes des grandes écoles et du deuxième et troisième cycles ; le BEPC se trouve aussi proche du CAP et du BEP que du baccalauréat.

Néanmoins, malgré cette transformation majeure de la structure de l'espace social, les deux ensembles identifiés au début de la période gardent des contours relativement proches,

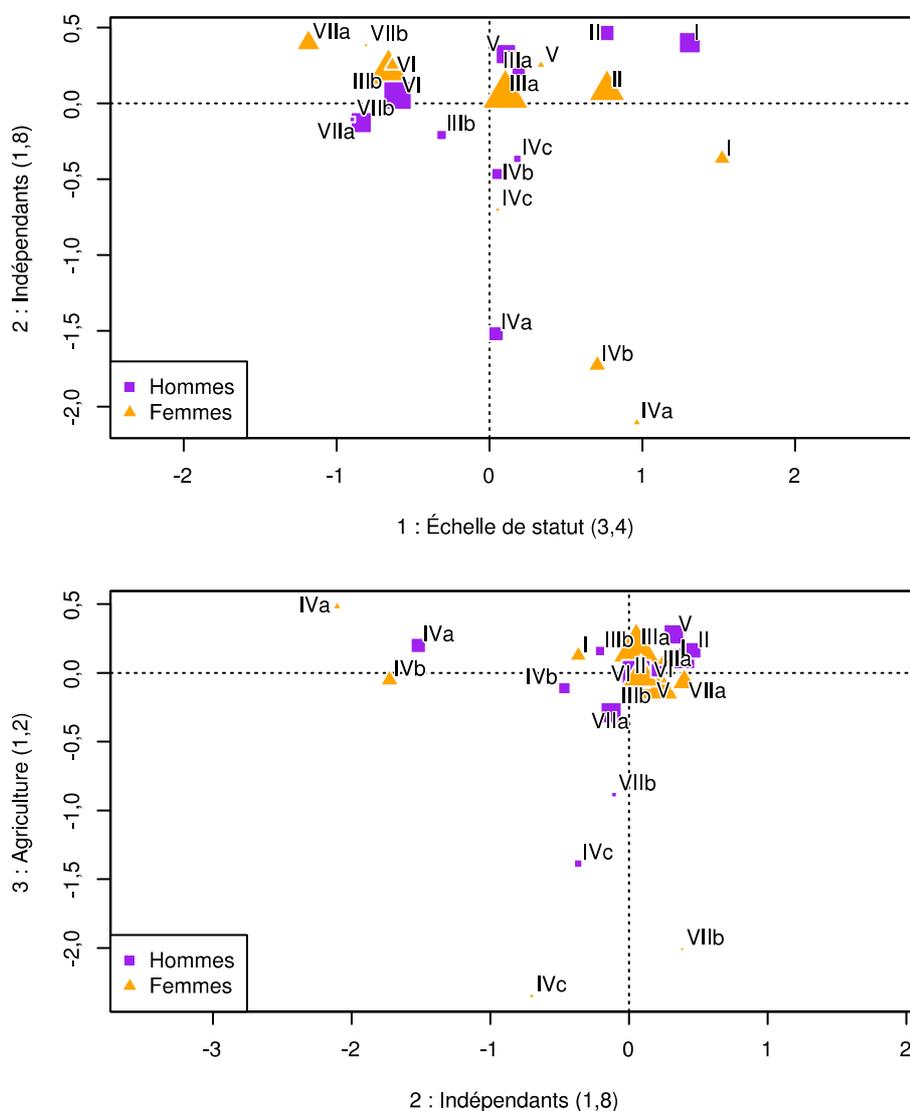
22 Du fait de l'utilisation de pondérations marginales, il n'est pas possible ici de comparer les positions entre années en valeur absolue : seule l'évolution des distances relatives entre groupes peut être commentée.

avec une frontière positionnée entre baccalauréats professionnel et général, faisant écho à la position centrale qu'occupe ce diplôme dans le système d'enseignement et dans le débat public. Au total, le seul bouleversement tient à la disparition de la spécificité des non diplômés. Pour le reste, la structure n'a pas radicalement changé, mais s'est simplement recomposée pour refléter une division accrue entre général et professionnel – quoique le fait que cette nouvelle distinction n'ait pas compensé complètement la disparition de la précédente constitue déjà un résultat notable.

4.3 L'homogamie de classe sociale : une baisse générale de l'association

Nous ne présenterons ici que les résultats portant sur la classe sociale des actifs, étant donné que l'inclusion des inactifs ne modifie pas les conclusions, mais réduit la netteté des dimensions obtenues. Le meilleur modèle selon le BIC est ici aussi le modèle RC-L à trois dimensions. La première dimension (Figure 2.6) correspond sans surprise à l'échelle de statut social, allant des ouvriers non qualifiés (VIIa) et des ouvriers agricoles (VIIb) aux classes supérieures (I). Une coupure nette est visible entre employés non qualifiés (IIIb) et qualifiés (IIIa), coupure que nous pouvons interpréter comme marquant la frontière entre classes populaires et classes moyennes et supérieures : les employés non qualifiés (IIIb) et les ouvriers (VI et VIIab) forment en effet un groupe relativement homogène en bas de la hiérarchie. Concernant son sommet, notons que les cadres et professions intellectuelles supérieures (I) sont situés nettement plus haut que ceux de niveau inférieur (II), groupe qui se détache lui-même assez clairement d'un ensemble central composé des employés qualifiés (IIIa), des techniciens (V) et des hommes indépendants (IVabc).

Figure 2.6 : Les dimensions du modèle RC(3)-L pour la classe sociale des actifs



Source : enquêtes Emploi (Insee), 1969-2011

Champ : couples cohabitants bi-actifs dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Une catégorie homme et une catégorie femme sont d'autant plus éloignées de l'origine et du même côté d'un axe (respectivement d'un côté différent) que leur combinaison est sur-représentée (resp. sous-représentée) parmi les configurations exogames par rapport à la situation de choix aléatoire du conjoint.

L'intensité de chaque dimension (*odds ratio* absolu moyen entre parenthèses, cf. p. 143) correspond à la moyenne sur la période. Une pondération par la moyenne des marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes.

Soulignons que la position des femmes indépendantes est clairement plus élevée dans la hiérarchie que celle des hommes de cette catégorie. Ce phénomène est probablement, au moins en partie, le résultat de différences dans la composition des deux groupes : les hommes indépendants tendent à exercer dans l'artisanat, le bâtiment et l'industrie, alors que les femmes tendent à exercer dans le tertiaire ; elles ont plus souvent travaillé comme employées avant de se mettre à leur compte (Fouquet, 2005). Néanmoins, les origines sociales des hommes et femmes indépendants sont similaires, et les femmes indépendantes ne sont que légèrement plus diplômées que leurs homologues masculins.

La deuxième dimension oppose les indépendants, et en particulier ceux employant des salariés (IVa), aux salariés. Notons que les femmes indépendantes sont fortement éloignées de l'origine, alors qu'à l'inverse toutes les autres catégories féminines, à l'exception des classes supérieures (I), sont situées de l'autre côté de l'axe : ceci est le signe d'une forte tendance des indépendantes à avoir un conjoint lui aussi indépendant, alors que les femmes salariées sont plus rarement que les hommes dans cette situation. Nous pouvons sans doute attribuer ce résultat au fait que les conjointes d'indépendants quittent plus facilement que les hommes leur emploi pour travailler avec leur conjoint (Bertaux-Wiame, 2004) ; la probabilité de se mettre à son compte au cours de l'année de mariage est deux fois plus élevée chez les femmes que chez les hommes (Toulemon, 1998, p. 40-42).

Seule exception à cette règle, les femmes cadres et professions intellectuelles supérieures (I) sont du même côté de la deuxième dimension que les hommes indépendants (IVabc), ce qui n'est pas le cas pour les hommes de ce groupe. De manière plus générale, on observe un chassé-croisé entre les positions des hommes et des femmes sur ces deux dimensions : alors qu'au sein des classes supérieures les femmes sont plus proches que les hommes des indépendants, le contraire est vrai au sein des classes populaires. Nous pouvons faire l'hypothèse que les femmes des classes supérieures quittent plus rarement leur emploi pour travailler avec leur conjoint indépendant que celles des classes populaires, et restent donc plus souvent classées comme salariées.

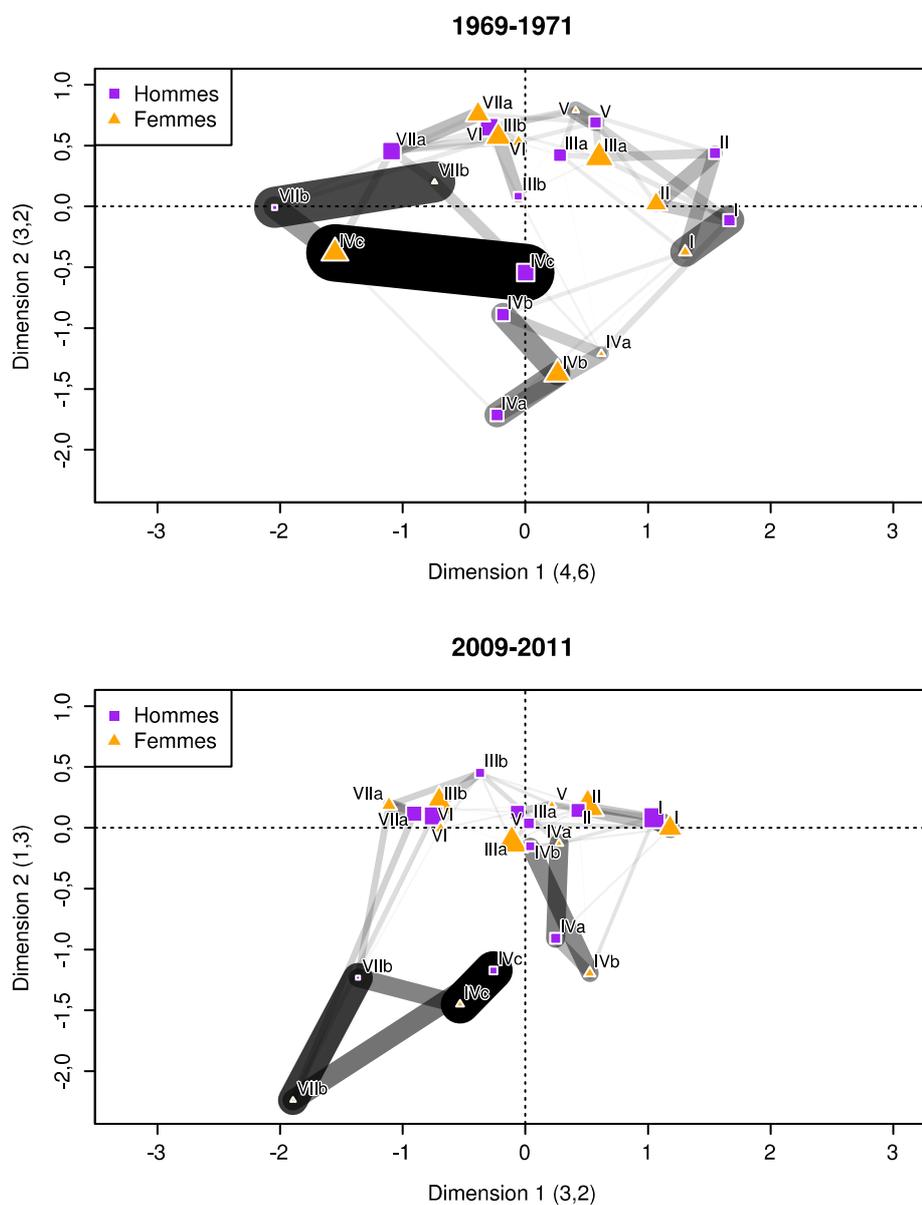
La troisième dimension oppose les agriculteurs (IVc) et les ouvriers agricoles (VIIb) à la plupart des autres groupes. Une fois encore, la position des femmes est plus prononcée sur cet axe que celle des hommes, un résultat que nous pouvons attribuer au même phénomène que pour les indépendants.

Les deux premières dimensions ont perdu en importance à partir des années 1980 (Figure 2.4, p. 163), et la troisième, qui était déjà assez faible, s'est encore légèrement affaiblie. Notons la présence d'un saut entre séries sur la première dimension en 1982, correspondant à l'échelle de statut social ; de nouveau, il apparaît que la refonte des PCS a amélioré l'identification des professions et in fine l'importance de l'échelle de statut mesurée à travers cette classification. Au total, le déclin global de l'homogamie de classe sociale observé avec le modèle UNIDIFF est clairement confirmé. Néanmoins, nous pouvons observer la quasi-disparition des deux dimensions qui correspondaient à des oppositions binaires plutôt qu'à des gradations (deuxième et troisième dimensions) : à la fin de la période, l'espace social s'ordonne presque exclusivement selon une seule dimension hiérarchique, l'échelle de statut social. Paradoxalement, alors que cette échelle a nettement perdu en importance, la perte de spécificité des indépendants, plus nette encore, a rendu la structure de classe plus unidimensionnelle et plus clairement ordonnée qu'auparavant.

La représentation transversale de la structure de l'espace social en début et en fin de période illustre ces tendances et permet de les préciser (Figure 2.7). La structure d'ensemble de l'homogamie est restée la même. On peut néanmoins identifier le resserrement de l'échelle de statut (première dimension, de même que dans les figures 2.44.2 et 2.6, p. 163 et 167), lié notamment au rapprochement des ouvriers non qualifiés (VIIa) du reste des classes populaires. Plus haut dans la hiérarchie, on peut noter que les cadres et professions intellectuelles de rang inférieur (II) se sont rapprochés de l'ensemble central constitué par les employés qualifiés (IIIa) et les techniciens et contremaîtres (V), alors que les cadres et professions intellectuelles supérieures (I) ont mieux résisté à ce mouvement de moyennisation. Comme pour l'homogamie de diplôme, la frontière marquant la limite des classes populaires n'a pas bougé : elle s'établit entre les employés non qualifiés (IIIb), d'un côté, et les employés qualifiés (IIIa) et les techniciens et contremaîtres (V) de l'autre. Paradoxalement au vu de l'effondrement politique de la classe ouvrière, les classes populaires forment un bloc plus homogène en 2009-2011 qu'en 1969-1971. Mais, dans cette définition restreinte qui exclut les employés qualifiés (IIIa), elles ont vu leur importance numérique clairement diminuer (*cf.* chapitre précédent) : elles ont ainsi cessé de constituer le centre de gravité de l'espace social, désormais occupé par un ensemble que l'on peut identifier aux classes moyennes, pour ne plus constituer que le bas de l'échelle sociale (Castel, 1999, repris dans son ouvrage de 2009).

La seconde évolution majeure, sur la seconde dimension, est constituée par la moindre spécificité des indépendants (IVab), déjà notée dans les figures 2.44.2 et 2.6 (p. 163 et 167). Du fait de la forte tendance à l'endogamie de ces groupes, il convient de rester prudent quant à l'analyse détaillée des positions des différentes catégories composant ce groupe, qui recouvrent des effectifs assez faibles dans les trois années d'enquête sur lesquelles reposent les modèles. La même remarque s'applique aux agriculteurs (IVc) et aux ouvriers agricoles (VIIb). Les résultats des modèles RC-L, qui mobilisent l'ensemble des enquêtes, sont de ce point de vue plus fiables.

Figure 2.7 : L'homogamie de classe sociale des actifs en 1969-1971 et en 2009-2011 selon le modèle RC(2)



Source : enquêtes Emploi (Insee)

Champ : couples cohabitants bi-actifs dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Une catégorie homme et une catégorie femme sont d'autant plus éloignées de l'origine et du même côté d'un axe (respectivement d'un côté différent) que leur combinaison est sur-représentée (resp. sous-représentée) parmi les configurations exogames par rapport à la situation de choix aléatoire du conjoint.

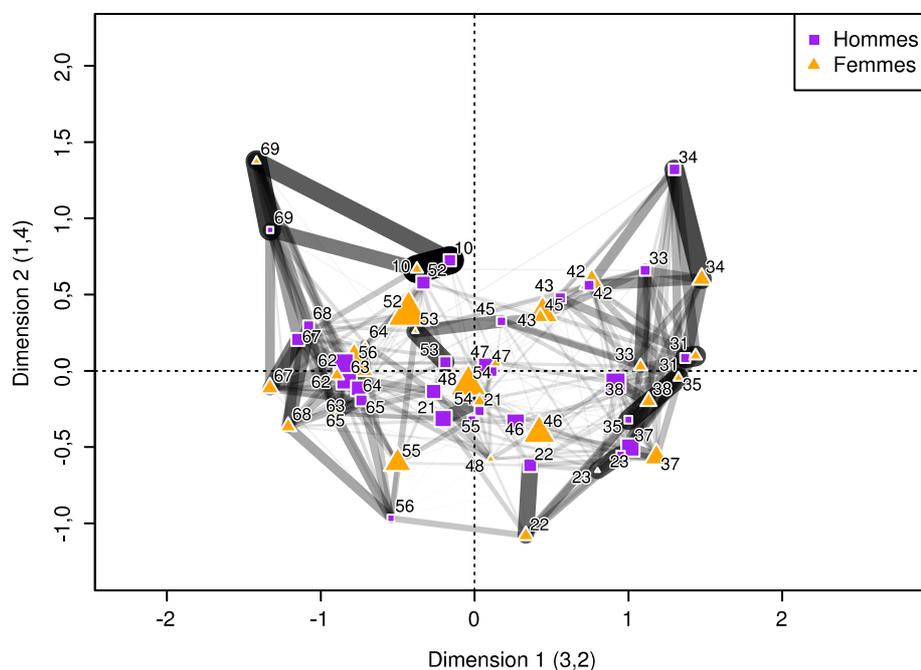
L'intensité de chaque dimension est indiquée par l'*odds ratio* absolu moyen entre parenthèses (*cf.* p. 143). Une pondération par les marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes. L'épaisseur et l'opacité des traits reliant les points hommes et femmes indiquent l'intensité de l'association positive entre les catégories correspondantes (*odds ratio* généralisé); les associations négatives ne sont pas représentées.

Le modèle portant sur les années 1969-1971 repose sur 17 641 observations et présente une déviance de 193 pour 53 degrés de liberté, avec un indice de dissimilarité de 3,4 %. Celui portant sur les années 2009-2011 repose sur 30 202 observations et présente une déviance de 206 pour 53 degrés de liberté, avec un indice de dissimilarité de 2,2 %.

Afin de confirmer ces résultats obtenus en recourant à la classification EGP, nous représentons ci-dessous l'espace social correspondant à l'homogamie observée en 2009-2011 à l'aide du niveau 3 de la nomenclature des PCS en 28 catégories (*cf.* Encadré 2.1)²³. L'image n'est pas radicalement bouleversée, mais cette nomenclature plus détaillée met en évidence certains points qu'EGP ne permettait pas de saisir (Figure 2.8). Tout d'abord, nous retrouvons bien un certain isolement des cadres et professions intellectuelles supérieures (3x dans les PCS ; I dans EGP), qui forment ici un réseau très dense d'associations entre les différentes catégories qui le composent ; ainsi qu'un ensemble opposé correspondant aux ouvriers (6x ; VI et VIIab) et aux employés des services directs aux particuliers (56, inclus dans IIIb dans EGP). Entre ces deux pôles, on observe un ensemble central composé du reste des employés (5x ; IIIab), des professions intermédiaires (4x ; réparties entre IIIa, II, V et IVab) et des petits indépendants (21 et 22 ; IVb pour la plupart).

23 On pourra trouver une analyse plus détaillée d'une table similaire dans les articles déjà cités (Bouchet-Valat, 2015b, 2015c).

Figure 2.8 : L'homogamie de PCS des actifs en 2009-2011 selon le modèle RC(2)



Source : enquêtes Emploi (Insee)

Champ : couples cohabitants bi-actifs dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Cf. Encadré 2.1 pour la description des codes.

Une catégorie homme et une catégorie femme sont d'autant plus éloignées de l'origine et du même côté d'un axe (respectivement d'un côté différent) que leur combinaison est sur-représentée (resp. sous-représentée) parmi les configurations exogames par rapport à la situation de choix aléatoire du conjoint.

L'intensité de chaque dimension est indiquée par l'*odds ratio* absolu moyen entre parenthèses (cf. p. 143). Une pondération par les marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes. L'épaisseur et l'opacité des traits reliant les points hommes et femmes indiquent l'intensité de l'association positive entre les catégories correspondantes (*odds ratio* généralisé) ; les associations négatives ne sont pas représentées.

Le modèle repose sur 30 195 observations et présente une déviance de 1 036 pour 548 degrés de liberté, avec un indice de dissimilarité de 5,9 %.

Encadré 2.2 : Niveau 3 de la nomenclature des PCS 2003

- 10 : Agriculteurs
- 21 : Artisans
- 22 : Commerçants et assimilés
- 23 : Chefs d'entreprise de 10 salariés ou plus
- 31 : Professions libérales
- 33 : Cadres de la fonction publique
- 34 : Professeurs, professions scientifiques
- 35 : Professions de l'information, des arts et des spectacles
- 37 : Cadres administratifs et commerciaux d'entreprise
- 38 : Ingénieurs et cadres techniques d'entreprise
- 42 : Professeurs des écoles, instituteurs et assimilés
- 43 : Professions intermédiaires de la santé et du travail social
- 44 : Clergé, religieux
- 45 : Professions intermédiaires administratives de la fonction publique
- 46 : Professions intermédiaires administratives et commerciales des entreprises
- 47 : Techniciens
- 48 : Contremaîtres, agents de maîtrise
- 52 : Employés civils et agents de service de la fonction publique
- 53 : Policiers et militaires
- 54 : Employés administratifs d'entreprise
- 55 : Employés de commerce
- 56 : Personnels des services directs aux particuliers
- 62 : Ouvriers qualifiés de type industriel
- 63 : Ouvriers qualifiés de type artisanal
- 64 : Chauffeurs
- 65 : Ouvriers qualifiés de la manutention, du magasinage et du transport
- 67 : Ouvriers non qualifiés de type industriel
- 68 : Ouvriers non qualifiés de type artisanal
- 69 : Ouvriers agricoles

(Seules les catégories d'actifs sont reprises ici. Les trois catégories d'agriculteurs ont été regroupées.)

Mais ces ensembles présentent une forte différenciation interne sur la seconde dimension, qui oppose les salariés du secteur public, agriculteurs et salariés agricoles (en haut) aux salariés du secteur privé et aux autres indépendants (en bas). Cette opposition s'exprime assez fortement au sein de toutes les PCS de niveau 1, à l'exception des ouvriers (6x) – chez qui ce critère n'intervient pas dans la définition des catégories. On observe ainsi que le lien entre classes moyennes et classes supérieures s'établit par deux circuits bien distincts, l'un passant par les professions intermédiaires du privé (46), l'autre par celles du public (42, 43 et 45). Ainsi, la prise en compte dans la nomenclature des PCS du critère du statut d'emploi – public ou privé – fait apparaître une distinction absente de la classification EGP, et qui a dans la période récente plus d'importance pour la description de l'homogamie que l'opposition entre salariés et indépendants du fait de la diminution de la taille du second groupe. Soulignons néanmoins que cette distinction n'est pas non plus retenue par le niveau 1 de la nomenclature des PCS : son absence ne révèle pas tant une lacune de la classification EGP, que le fait qu'elle ne possède pas de niveau plus détaillé que celui que nous avons utilisé ici²⁴.

4.4 L'homogamie de classe sociale d'origine : une plus grande stabilité

La classe sociale des pères des conjoints n'est disponible que depuis 1982 et a été codée de manière moins détaillée : il n'est pas étonnant qu'un modèle RC-L à deux dimensions seulement soit suffisant selon le BIC pour décrire correctement les données²⁵. Ces dimensions prennent leur signification lorsqu'elles sont envisagées ensemble, plutôt que séparément comme nous l'avons fait jusqu'ici.

Ici encore, nous pouvons identifier la hiérarchie des professions salariées partant du coin inférieur gauche et s'achevant dans le coin supérieur droit du plan (Figure 2.9). Les indépendants (IVab) constituent un groupe intermédiaire entre les cadres et professions intellectuelles (I et II) et les agriculteurs (IVc), alors que les ouvriers agricoles (VIIb) occupent une position équivalente entre ces derniers et le reste des classes populaires, quoique avec un statut inférieur. La perte d'importance des deux dimensions, bien que

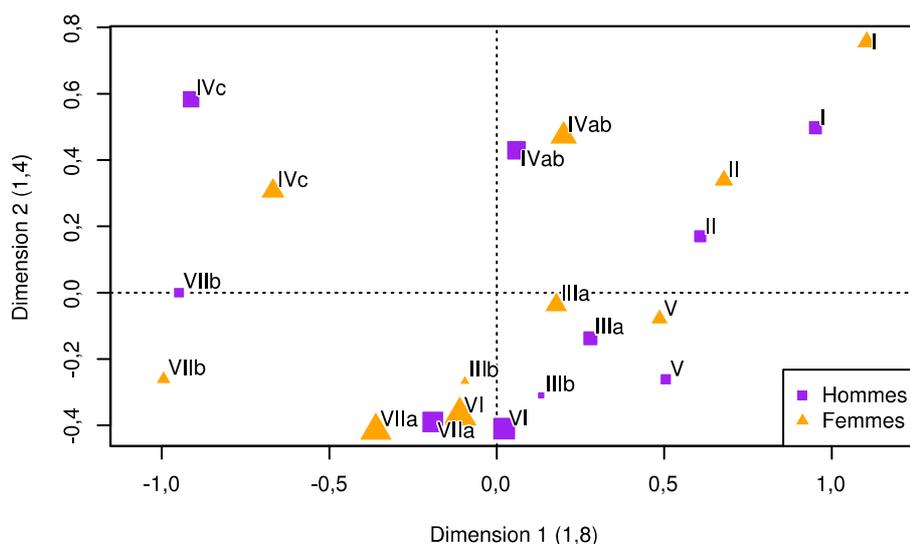
24 La distinction public/privé, si elle prend des formes différentes selon les pays, semble un facteur de stratification assez commun pour justifier sa prise en compte dans une nomenclature internationale (Hugrée, Pénissat et Spire, 2015).

25 Ces deux dimensions sont dans une large mesure similaires à celles décrivant l'association entre les classes sociales des conjoints dans un modèle RC-L à deux dimensions (non repris ici).

statistiquement significative, est bien moins prononcée que pour les deux autres composantes de l'homogamie (Figure 2.4, p. 163). L'affaiblissement de l'association est donc plus marqué pour les configurations endogames que pour les configurations exogames²⁶.

Les représentations transversales fournies par les modèles RC confirment ce résultat (Figure 2.10) : la stabilité d'ensemble est frappante. Néanmoins, on note un rapprochement significatif des cadres et professions intellectuelles supérieures (I) des classes moyennes. Cette évolution, qui semble expliquer le léger déclin de l'association parmi les configurations exogames, fait écho à la diminution de l'endogamie relative de ce groupe déjà relevée plus haut.

Figure 2.9 : Les dimensions du modèle RC(2)-L pour la classe sociale d'origine



Source : enquêtes Emploi (Insee), 1982-2011

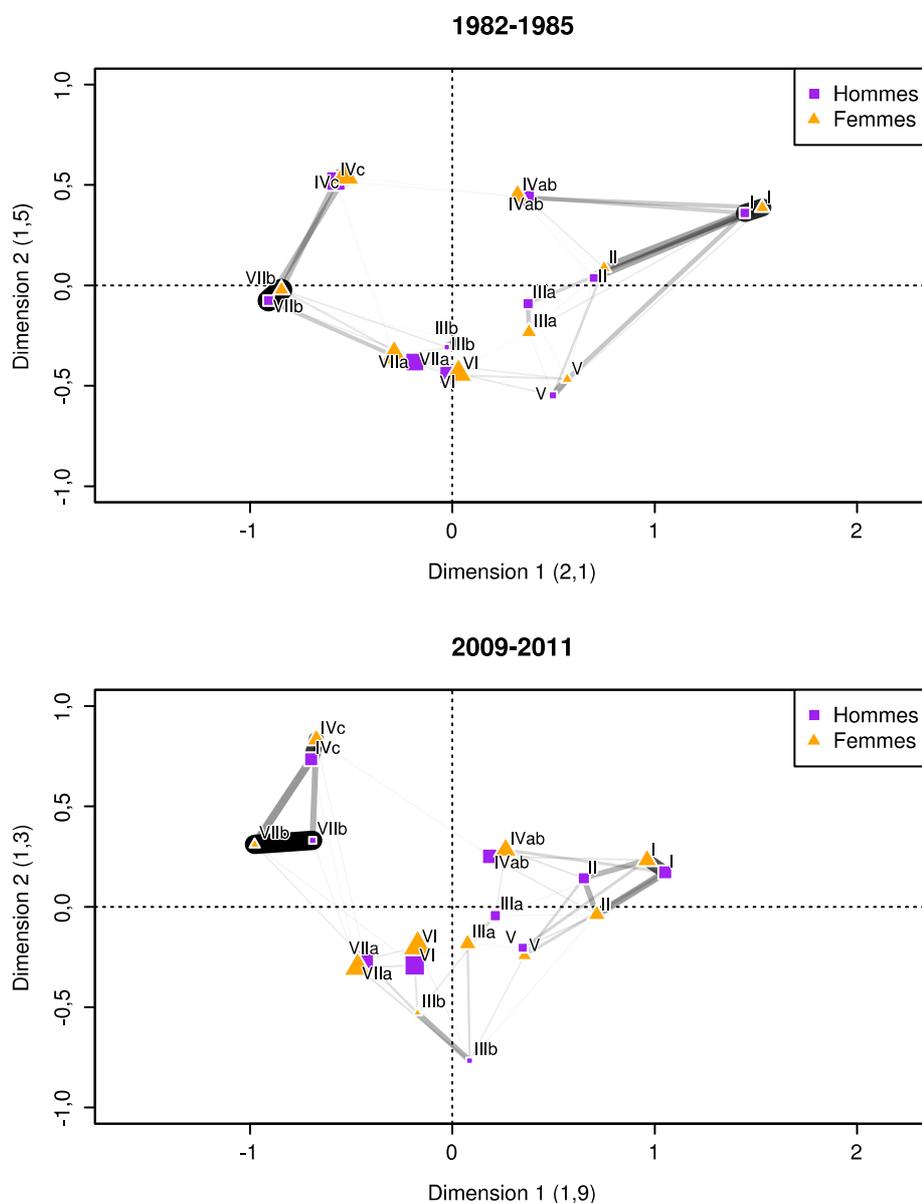
Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Une catégorie homme et une catégorie femme sont d'autant plus éloignées de l'origine et du même côté d'un axe (respectivement d'un côté différent) que leur combinaison est sur-représentée (resp. sous-représentée) parmi les configurations exogames par rapport à la situation de choix aléatoire du conjoint.

L'intensité de chaque dimension (*odds ratio* absolu moyen entre parenthèses, cf. p. 143) correspond à la moyenne sur la période. Une pondération par la moyenne des marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes.

26 Ce qui se vérifie par la comparaison de modèles UNIDIFF incluant et n'incluant pas la diagonale de la table.

Figure 2.10 : L'homogamie de classe sociale d'origine en 1982-1985 et en 2009-2011 selon le modèle RC(2)



Source : enquêtes Emploi (Insee)

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans

Une catégorie homme et une catégorie femme sont d'autant plus éloignées de l'origine et du même côté d'un axe (respectivement d'un côté différent) que leur combinaison est sur-représentée (resp. sous-représentée) parmi les configurations exogames par rapport à la situation de choix aléatoire du conjoint.

L'intensité de chaque dimension est indiquée par l'*odds ratio* absolu moyen entre parenthèses (cf. p. 143). Une pondération par les marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes. L'épaisseur et l'opacité des traits reliant les points hommes et femmes indiquent l'intensité de

l'association positive entre les catégories correspondantes (*odds ratio* généralisé); les associations négatives ne sont pas représentées.

Le modèle portant sur les années 1982-1985 repose sur 35 884 observations et présente une déviance de 53 pour 39 degrés de liberté, avec un indice de dissimilarité de 1,4 %. Celui portant sur les années 2009-2011 repose sur 37 759 observations et présente une déviance de 47 pour 39 degrés de liberté, avec un indice de dissimilarité de 1,4 %.

5 Conclusion

L'homogamie en termes de diplôme, de classe sociale et de classe sociale d'origine a clairement décliné en France au cours des quarante dernières années. Parmi les couples cohabitants, mariés ou non, dans lesquels l'un des conjoints est âgé de 30 à 59 ans, la proportion de couples endogames (ceux dans lesquels, respectivement, les deux conjoints possèdent le même diplôme, appartiennent à la même classe sociale ou sont issus de la même classe sociale) a baissé entre 1969 et 2011 de 30 à 50 %, selon la composante retenue. Dans le cas du diplôme, ces changements sont en partie un effet mécanique de la massification scolaire, qui a fortement réduit la probabilité qu'un homme et une femme pris au hasard aient exactement le même diplôme. Cependant, même en contrôlant ces évolutions structurelles, nous observons une division par près de 2 des *odds ratios* mesurant l'intensité de l'homogamie relative du point de vue du diplôme, et de celui de la classe sociale pour les couples bi-actifs ; et une réduction d'un tiers du point de vue de la classe sociale en incluant les inactifs, et de la classe sociale d'origine. Sur l'ensemble de la période, la ressemblance des conjoints en termes de classe sociale d'origine est inférieure d'un tiers à celle en termes de diplôme et de classe sociale ; et quoique ces dernières soient d'intensité comparable, il semble que dans les années récentes l'homogamie de classe sociale tende à devenir plus faible que celle en termes de diplôme.

D'autre part, nous avons observé une inversion de la différence moyenne entre le diplôme de l'homme et celui de la femme au sein des couples : alors qu'en 1969 les hommes étaient plus diplômés que leurs conjointes (hypergamie féminine), l'inverse est vrai depuis l'an 2000. Ce phénomène persiste lorsque l'on tient compte de l'évolution de la structure de la population par diplôme et par sexe, ce qui peut être considéré comme le signe que les comportements de choix du conjoint, et donc potentiellement les normes sociales, se sont

modifiés face à la nouvelle situation dans laquelle les femmes sont globalement plus diplômées que les hommes.

L'analyse détaillée des évolutions donne plus de sens à ces tendances globales. Le déclin de l'homogamie de diplôme résulte de plusieurs tendances divergentes affectant différentes dimensions de la stratification en termes de diplôme. Nous observons en fait à travers l'homogamie les conséquences de la massification scolaire. D'un côté, le groupe des non diplômés a progressivement perdu de sa spécificité et est devenu plus perméable aux autres groupes. De l'autre, le choix du conjoint s'ordonne toujours autant, voire un peu plus, selon l'échelle des diplômes. L'endogamie a diminué parmi tous les groupes, la décreue étant plus nette pour les non diplômés et les diplômés de l'Université, qui restent néanmoins parmi les plus endogames. Mais les diplômés des grandes écoles et les titulaires du Certificat d'études – aujourd'hui disparu –, présentent une endogamie encore plus forte, et sont les seuls groupes à l'avoir renforcée. Ce résultat offre une nouvelle illustration de la forte spécificité des grandes écoles au sein de l'enseignement supérieur français.

Contrairement à ce que nous observons concernant l'homogamie de diplôme, l'affaiblissement de l'homogamie de classe sociale, qui est intervenu depuis le milieu des années 1980, est le résultat de tendances convergentes à la baisse. Trois dimensions de la stratification selon la classe sociale ont été identifiées : l'échelle de statut des professions ; l'opposition entre indépendants et salariés ; l'opposition entre les professions agricoles et les autres. Toutes ces dimensions ont décliné, mais seule la première conserve une véritable importance aujourd'hui, si bien que l'espace des classes sociales s'est largement unifié autour d'une unique dimension structurante. L'homogamie selon la classe sociale d'origine des conjoints suit un scénario très similaire, mais la baisse concerne principalement l'endogamie, et semble ralentir dans la période récente.

La disponibilité d'enquêtes annuelles nous a permis d'évaluer l'effet des changements intervenus entre séries, une possibilité rarement ouverte aux chercheurs. Malgré quelques petites ruptures, la cohérence de la série des enquêtes Emploi est très bonne. Même si nous pouvions craindre une diminution au cours du temps de l'adéquation entre les catégories socioprofessionnelles (et la classification EGP fondée sur elles) et la réalité qu'elles décrivent, la refonte majeure introduite en 1982 n'a causé qu'un changement relativement faible du niveau observé de l'homogamie ; la révision réalisée par l'Insee en 2003 n'est

même pas visible dans nos mesures. Le déclin de l'homogamie de classe sociale en termes absolus et relatifs semble donc robuste.

Il semble intéressant de mettre nos résultats en relation avec des travaux portant sur des questions proches. La plus forte tendance à l'entre-soi des classes supérieures, par rapport aux classes moyennes et populaires, que nous observons, a aussi été signalée dans des travaux portant sur la ségrégation socio-spatiale (Maurin, 2004 ; Préteceille, 2006 ; Godechot, 2013), les deux derniers concluant à son renforcement pour les fractions économiquement dominantes, quand les autres groupes auraient plutôt tendance à voir leur ségrégation diminuer légèrement depuis les années 1990. Ce phénomène s'observe aussi du point de vue des inégalités de revenus, qui ont surtout augmenté au sommet de la distribution (Piketty, 2001, 2003 ; Landais, 2007 ; Amar, 2010 ; Solard, 2010 ; Godechot, 2012), et, parmi les salariés, presque exclusivement ceux d'Île-de-France (Godechot, 2013). Les catégories retenues dans la présente étude sont bien plus larges que le centile supérieur de la distribution des revenus : notre analyse décrit des tendances affectant la majorité de la société, mais n'est pas adaptée à l'étude de petits groupes aux comportements extrêmes. Nous pouvons néanmoins considérer l'augmentation de l'endogamie des diplômés des grandes écoles comme le signe, dans le domaine du choix du conjoint, d'un tel renforcement du séparatisme des classes supérieures, séparatisme qui n'a rien de « généralisé » (Maurin, 2004) à toute la société, tant il contraste avec l'évolution observée pour la quasi-totalité des autres groupes.

Il paraît difficile de ne pas mettre en relation l'affaiblissement de l'homogamie que nous observons avec la baisse du sentiment d'appartenance à une classe sociale et la perte de visibilité de cette notion dans le débat public. En effet, quel que soit le niveau des inégalités entre groupes sociaux, l'intensité des relations entre ces groupes est sans aucun doute déterminante dans la conception que se font les individus du monde social. Il ne semble donc pas que la « spirale des classes sociales » décrite par L. Chauvel (2001, 2006) soit aujourd'hui sur le point de virer, suite à une augmentation des inégalités, dans la direction d'un renforcement de la conscience de classe. Au contraire, nous avons observé que l'endogamie, indice de la force des identités de groupe, a nettement décliné. À la place d'oppositions entre groupes distincts, les positions sociales s'inscrivent aujourd'hui plus

qu'hier dans un gradient (Castel, 1999, 2009 ; Bouffartigue, 1999, 2001)²⁷, qui rappelle la description que donne Simmel (1897, p. 102) d'une société « *dirigée* » par la classe moyenne : « *Une société de ce genre a pour caractère distinctif la continuité. Elle n'implique, en effet, ni une égalité absolue entre les individus, ni la division du groupe en deux parties radicalement hétérogènes, l'une supérieure, l'autre inférieure* ». Ainsi, plutôt qu'à un retour des classes sociales, nous semblons assister à une unification lente mais régulière de la société française, qui contraste avec un mouvement de repli décelable à son extrême sommet.

Le déclin de la conscience de classe ne semble donc pas tenir à un simple décalage entre la réalité des inégalités et leur perception : il est plus fondamentalement inscrit dans les relations entre groupes sociaux, par le brouillage de leurs frontières et leur moindre stabilité. On peut penser que la massification scolaire, l'expansion du groupe des catégories supérieures et des employés et l'augmentation de la mobilité sociale ont joué un rôle important dans cette évolution, du fait de l'entrecroisement des appartenances (Simmel, [1908] 1999 ; Blau et Schwartz, 1984) suivant le mécanisme décrit en introduction.

À l'encontre de cette conclusion, on peut avancer une critique portant sur la continuité temporelle des catégories utilisées. Il est entendu que les cadres et les ouvriers de 2011 sont fort différents de leurs homologues de 1969 ; dès lors, doit-on envisager que les véritables clivages sociaux se soient simplement déplacés vers d'autres frontières qui ne correspondraient plus aux catégories classiques retenues ici ? Privilégiant une logique positionnelle (Bourdieu, 1966), on pourrait considérer que les classes supérieures forment un groupe de taille relativement stable à travers le temps, et qu'en conséquence, suite à l'expansion de ces groupes, les véritables distinctions passent désormais à l'intérieur de la catégorie des diplômés du supérieur ou de celle des cadres.

Une telle lecture trouve un certain appui dans l'augmentation de l'endogamie des diplômés des grandes écoles que nous avons relevée : on peut donc penser qu'une définition plus fine et plus restrictive des classes supérieures permettrait de mettre en évidence, à partir de la profession, le même mouvement de repli. Du point de vue de l'éducation également, il ne semble pas inenvisageable qu'une distinction entre diplômés du supérieur suivant le domaine d'études donnerait à voir une tendance temporelle différente pour les

27 Ce résultat, mis en évidence plus haut par la perte d'importance des dimensions secondaires de la stratification sociale, se vérifie plus formellement par le fait qu'un modèle RC-L à une seule dimension et sans paramètres spécifiques à la diagonale décrit mieux les données à la fin qu'au début de la période.

filières les plus socialement valorisées (droit, médecine, économie et gestion notamment). Néanmoins, si de tels prolongements permettaient de préciser, et sans doute de conforter, notre constat d'un mouvement de repli au sommet de l'échelle sociale, ils ne semblent pas susceptibles de remettre fondamentalement en cause celui d'une moindre détermination sociale du choix du conjoint dans l'ensemble de la société, tant les groupes concernés sont – du fait de la définition même de l'approche positionnelle – de taille restreinte.

Plus sérieuse nous apparaît une seconde objection, fondée celle-là sur une logique non plus de position, mais de condition : du fait de la relative perte de valeur des diplômes et de la déstabilisation du statut de cadre, il est possible que les dimensions de la stratification retenues ici reflètent de moins en moins bien les conditions de vie des individus. En effet, l'augmentation de la fluidité sociale et la baisse des inégalités scolaires à l'œuvre en France au cours des dernières décennies contrastent avec l'accroissement des inégalités économiques, du chômage et de la précarité de l'emploi. Les clivages pertinents entre groupes sociaux pourraient ainsi être devenus désormais plus transversaux aux groupes sociaux définis par le diplôme et la profession.

Précisons dès ici que cette éventualité nous semble peu probable. D'une part, comme on l'a noté, l'augmentation des inégalités de revenus et de patrimoine concerne presque exclusivement l'extrême sommet de la distribution : nous retrouvons en fait l'argument déjà discuté ci-dessus. D'autre part, le chômage et la précarité frappent avant tout les classes populaires (ouvriers, employés, moins diplômés), venant plus renforcer qu'atténuer ces dimensions de la stratification sociale. Cependant, cette objection mérite à notre sens d'être mise à l'épreuve, et l'étude du lien entre les situations des conjoints du point de vue économique constitue une entrée alternative riche en enseignements pour la compréhension de l'homogamie. C'est pourquoi le prochain chapitre, concentrant cette investigation sur un aspect de la question, traitera de l'évolution du lien entre les positions des conjoints sur le marché du travail.

CHAPITRE 3

LA SITUATION SUR LE MARCHÉ DU TRAVAIL : EFFET ÉGALISATEUR DE L'ACTIVITÉ FÉMININE MAIS STABILITÉ DU LIEN ENTRE STATUTS D'EMPLOI

Dans les deux chapitres précédents, nous avons mis en évidence un affaiblissement notable de l'homogamie en termes absolus comme relatifs, du point de vue de dimensions essentielles de la stratification sociale : l'éducation, la classe sociale et l'origine sociale. Ce résultat semble indiquer l'existence d'un mouvement vers une moins grande proximité entre les caractéristiques sociales des conjoints. Néanmoins, comme nous l'avons discuté en conclusion, on ne peut exclure qu'il signale plutôt une perte d'influence des critères de définition de la position sociale traditionnellement utilisés par la sociologie, du fait de l'augmentation de l'hétérogénéité interne aux groupes identifiés par ces variables classiques.

De fait, alors que l'éducation a connu une massification sans précédent et que la structure par classe sociale a bénéficié d'une relative croissance des positions supérieures, on assiste depuis plusieurs décennies à une augmentation des inégalités économiques due à la forte croissance des revenus dans le haut de la structure sociale (Piketty, 2001, 2003 ; Landais, 2007 ; Amar, 2010 ; Solard, 2010 ; Godechot, 2012, 2013), et à une précarité accrue dans sa partie inférieure (Paugam, 1991 ; Castel, 1995). En s'en tenant au constat de l'existence d'un mouvement de moindre détermination du choix du conjoint du point de vue des variables classiques de la sociologie, on s'expose par conséquent au risque de passer à côté d'une reconstitution potentielle des inégalités sociales sous d'autres aspects. Néanmoins, ce risque n'est qu'une éventualité : les inégalités ne se sont pas nécessairement développées au point de contrer le mouvement d'ouverture déjà mis en évidence. À titre d'exemple, si l'on excepte le sommet de la distribution, c'est plutôt une légère baisse des

inégalités de revenus qui prend place depuis une vingtaine d'années (Pujol et Tomasini, 2009 ; Verdugo, Fraisse et Horny, 2012 ; et Charnoz, Coudin et Gaini, 2013 concernant les salaires des hommes à temps complet dans le secteur privé), à rebours du discours dominant sur la question. De même, il est loin d'être certain que l'effritement du salariat puisse véritablement donner lieu à la création de nouvelles lignes de démarcation entre groupes sociaux, plutôt qu'à une précarisation plus diffuse.

Afin de tenter de répondre à ces interrogations, nous traiterons dans ce chapitre de l'association entre les positions des conjoints sur le marché du travail. Toujours à partir des enquêtes Emploi, nous nous intéresserons tout d'abord à l'évolution depuis 1982 (ou parfois 1990), du lien entre les salaires des conjoints et son effet sur les inégalités entre les couples, avant de porter notre attention sur différents indicateurs de précarité de l'emploi depuis 1975 (ou 1982).

La première entrée, par les salaires, visera notamment à évaluer l'effet du développement massif de l'emploi des femmes sur les inégalités entre couples. Nous suivons en cela une veine de travaux portant principalement sur le cas des États-Unis, pour lequel il a été montré que cette évolution a eu un effet modérateur sur la hausse des inégalités, mais que l'augmentation de la proximité des conjoints en termes de salaires a eu l'effet inverse (Cancian et Reed, 1998 ; Hyslop, 2001 ; Schwartz, 2010). Nous laissons ici de côté la question de l'homogamie liée au patrimoine, ayant déjà fait l'objet d'une étude transversale récente en France (Frémeaux, 2014), qui exige de recourir à des sources de données de nature différente. En outre, si la restriction aux seuls salaires revient à laisser échapper une partie importante de l'augmentation des inégalités de revenus (Piketty, 2001, 2003 ; Landais, 2007), cette approche garde néanmoins tout son intérêt pour étudier particulièrement les effets de l'augmentation de l'activité des femmes sur le lien entre les revenus des conjoints, ainsi que sur l'évolution des rôles de genre.

La seconde entrée, par la précarité de l'emploi, aura pour objet la question de l'éventuelle apparition de nouveaux clivages de la société française, qui, tout en touchant particulièrement les catégories populaires, traverseraient en partie les classes sociales. Il s'agira d'évaluer si on assiste à la constitution d'un groupe social doté d'une certaine cohérence, désigné alternativement sous le nom d'*under-class*, de « monde des exclus » (Dubet, 2001), de « précarariat » (Castel, 2009 ; Standing, 2011) ou encore d'« exilés du noyau

dur du salariat » (Eckert, 2010). L'apparition d'un tel groupe serait alors la conséquence d'une concentration de la précarité, du fait de la forte segmentation du marché du travail français, sur certains secteurs d'activité, certaines catégories socioprofessionnelles, certains profils sociaux, ou encore certaines zones géographiques. Cependant, le passage de cette situation objective à la constitution d'un groupe social – observée à travers les alliances conjugales – paraît relativement improbable, du fait de l'importante hétérogénéité de la catégorie des précaires (Vultur, 2010), ainsi que des mécanismes d'assistance qui vont à l'encontre de l'établissement de solidarités ou d'une conscience commune (Paugam, 1991, chap. 4 et 5, 2000, chap. 9 et conclusion). En effet, on sait au moins depuis les travaux de Georg Simmel ([1907] 1998) que la catégorie des « pauvres » ou des « assistés » ne se définit pas par son homogénéité interne, mais par sa position dans la société. On peut en outre souligner la relative instabilité temporelle de ce groupe, constitué en grande partie de jeunes en insertion difficile sur le marché du travail (Cabannes, 2014)¹, qui finiront pour la plupart par trouver un emploi stable – quoique certains oscilleront durablement entre contrats à durée limitée et chômage (Castel, 1995, p. 662 ; Eckert et Mora, 2008).

À la différence des composantes de l'homogamie déjà étudiées, dont on peut considérer qu'elles sont fixées avant la rencontre des conjoints, soit complètement (dans le cas de l'origine sociale), soit en grande partie (dans le cas du diplôme), ou du moins dans une certaine mesure (dans le cas de la classe sociale), la position des individus sur le marché du travail que nous étudierons ici résulte très largement de décisions coordonnées entre les conjoints qui dépendent de la situation économique du couple et des rôles de genre (Verbakel, 2008, chap. 3). Ces variables doivent ainsi être considérées comme fondamentalement endogènes. Nous ne parlerons donc pas d'homogamie de salaire, mais d'association entre les salaires des conjoints. Cette expression neutre recouvre dans notre lecture à la fois les caractéristiques des conjoints antérieures à la mise en couple (niveau de qualification notamment), et des décisions en termes d'activité et de carrière prises plus tard. En particulier, dans le modèle fortement genré des années 1960 qui persiste encore de manière atténuée aujourd'hui, le taux d'emploi des femmes varie en fonction des revenus de leur conjoint, ainsi que des revenus qu'elles peuvent espérer obtenir sur le marché du travail.

1 Mais aussi secondairement de salariés âgés éprouvant des difficultés à retrouver un emploi après un licenciement. Ces deux catégories ont en commun de se situer aux deux extrêmes de la vie active.

Néanmoins, si l'interdépendance des décisions prises par les conjoints interdit de considérer leur proximité en termes de salaires comme un pur symptôme de la plus ou moins grande distance entre groupes sociaux, ces décisions n'en ont pas moins des conséquences importantes sur le niveau des inégalités entre couples et entre hommes et femmes. Ainsi, alors que nous avons privilégié une lecture de l'homogamie en tant que symptôme de la structure sociale, nous porterons également dans ce chapitre notre attention sur ses conséquences, suivant en cela un courant de recherches relativement récent (Schwartz, 2013). Nous traiterons aussi plus longuement la question de l'hypergamie féminine, c'est-à-dire la question centrale de l'infériorité du salaire des femmes dans le couple, amorçant ainsi des analyses que nous développerons plus en détail dans la deuxième partie.

Nous commençons par développer la construction des différentes variables économiques retenues ici à partir des enquêtes Emploi, déjà présentées dans les chapitres précédents. Nous montrons ensuite, en étudiant l'évolution des salaires des hommes et des femmes en couple cohabitant, que la croissance de l'activité des femmes a entraîné une diminution significative des inégalités de salaire entre couples, du fait notamment de la stabilité de la corrélation entre salaires des conjoints, elle-même liée à la stabilité de la distribution sociale de l'inactivité féminine. Grâce à une méthode de décomposition fondée sur différents scénarios contrefactuels, nous mettons en évidence le rôle assez limité de l'évolution de l'association entre salaires des conjoints dans cet affaiblissement des inégalités. En effet, des analyses plus détaillées montrent que ce lien a légèrement perdu en intensité au cours du temps parmi les couples bi-actifs ; en revanche, l'hypergamie féminine en termes de salaire ne semble pas avoir diminué au-delà de ce qu'implique l'augmentation du taux d'activité féminine et des salaires relatifs des femmes. Finalement, nous confirmons le résultat selon lequel les inégalités sur le marché du travail ne se sont pas substituées aux dimensions classiques de la stratification sociale en mettant en évidence une grande stabilité de la ressemblance des conjoints en termes de précarité du statut d'emploi.

1 Les données : salaire et précarité de l'emploi dans les enquêtes Emploi

Nous recourons ici encore à la série des enquêtes Emploi, déjà décrite au chapitre 1, en nous concentrant toujours sur les couples cohabitants dans lesquels l'un des deux conjoints a entre 30 et 59 ans. Comme pour l'étude d'autres dimensions de l'homogamie, ces enquêtes offrent l'intérêt de fournir des données annuelles fondées sur des échantillons de taille importante, et de renseigner sur le long terme de manière relativement homogène des informations assez détaillées concernant le salaire, le statut d'emploi et le chômage. En revanche, les enquêtes ne contiennent avant 2003 aucune information concernant les revenus autres que salariaux, et ces derniers sont fournis sur une base purement déclarative². L'étude des salaires des conjoints nous semble néanmoins déjà assez riche en elle-même pour justifier un travail à partir de ces données.

La période d'étude s'ouvre ici, selon les analyses, en 1975, en 1982 ou en 1990. En effet, la mesure du chômage ne s'est stabilisée qu'à partir de la première date, et le salaire des individus et les indicateurs concernant la précarité de l'emploi ne sont disponibles qu'à partir de la seconde. En outre, le salaire était déclaré par tranches entre 1982 et 1989, et leurs bornes (19 au total, avec une amplitude minimale de 500 francs) n'étaient pas revalorisées d'une année sur l'autre pour tenir compte de l'augmentation des salaires : aussi, ces tranches n'autorisent pas une mesure fiable de l'évolution temporelle des inégalités. Nous retenons donc uniquement les enquêtes postérieures à 1990 dès que notre attention se porte sur cette question³.

1.1 Le salaire

Le questionnaire des enquêtes Emploi interroge les individus sur le montant de leur salaire mensuel le mois précédant l'enquête, ainsi que sur les éventuels compléments touchés annuellement (primes, 13^e mois...). Les enquêtés se voient proposer la possibilité de

2 Un recours aux enquêtes Revenus fiscaux et sociaux, associées aux enquêtes Emploi depuis 1996, permettrait à la fois de s'assurer de la conformité des salaires avec les informations déclarées à l'administration fiscale, et d'élargir l'étude aux autres types de revenus, au prix d'un raccourcissement de la période étudiée.

3 Il serait sans doute possible de contourner cette limite des données en imputant un salaire continu à partir des tranches et des autres variables disponibles dans la base, tout en se fondant sur les données des séries ultérieures pour contrôler la qualité de l'imputation ainsi réalisée. Nous n'avons pas expérimenté cette possibilité.

ne pas répondre, ou bien d'indiquer une tranche de salaire plutôt qu'un montant exact, auxquels cas l'Insee impute un montant à partir des autres variables disponibles, à l'aide d'un modèle dont on ne connaît malheureusement pas la spécification. Avant 2003, seul le salaire net était demandé ; depuis cette date, les interrogés peuvent choisir de déclarer le montant brut, qui est transformé en montant net *a posteriori* à l'aide de tables de passage.

Depuis 2003, les indépendants sont aussi interrogés sur leurs revenus professionnels. Nous avons choisi d'ignorer cette information par cohérence avec les séries précédentes, mais aussi à cause de la difficulté à individualiser un tel revenu, dont il est précisé explicitement qu'il est touché par « *vous et votre famille* ». Les couples comprenant un indépendant sont donc exclus de l'échantillon.

Les déclarations portant sur le mois précédant l'enquête, le questionnaire prévoit que seuls les actifs occupés au moment de l'enquête puissent déclarer un salaire. Nous attribuons systématiquement un salaire nul aux chômeurs et aux inactifs (moins de 3 % d'entre eux déclarent un salaire), de manière à clarifier la terminologie en considérant qu'une personne en emploi est une personne déclarant un salaire⁴. En revanche, nous ne réalisons aucune distinction entre salariés à temps complet et à temps partiel. Il ne s'agit pas pour nous d'étudier le salaire horaire, mais bien le salaire effectivement touché, résultat aussi bien de décisions d'activité plus ou moins choisies ou subies⁵ (et fréquemment liées à des situations familiales) que d'une position plus ou moins favorisée en termes de salaire horaire.

De même, à la suite de plusieurs travaux, nous n'ajustons pas les salaires du couple au nombre de personnes vivant dans le ménage de manière à approcher son niveau de vie. En effet, nous ne disposons pas de l'ensemble des revenus du ménage, et nous ne nous intéressons pas aux revenus de ses autres membres. Un tel ajustement, même s'il permettrait par exemple de mieux appréhender l'effet des évolutions étudiées sur le taux de pauvreté, rendrait les mécanismes à l'œuvre plus difficiles à interpréter, quand notre attention porte sur la relation entre les salaires des conjoints.

4 Dans la suite, nous désignerons parfois par simplicité comme « couples bi-actifs » les couples dans lesquels les deux conjoints touchent un salaire ; de même, sans autre précision, « activité féminine » fera référence au fait que la femme touche un salaire.

5 Suivant Monique Meron et Margaret Maruani (2012, chap. 2), nous considérons ici que la question du « choix » ne revêt que peu de signification, l'essentiel étant d'identifier les différents contextes et origines de l'emploi à temps partiel.

Du fait des importantes fluctuations liées à l'échantillonnage au sommet de la distribution des salaires, qui rendaient certains indicateurs instables d'une année sur l'autre, nous avons ramené pour chaque année au 995^e millime les salaires qui dépassaient ce niveau (soit 0,5 % des cas). Les effectifs de nos enquêtes ne nous permettent pas d'étudier précisément ce groupe (représentant environ 40 individus par an) d'une année sur l'autre.

Finalement, les salaires étant déclarés en francs jusqu'en 2002, nous les transformons en euros, avant de les corriger de l'inflation à l'aide de l'indice des prix à la consommation calculé par l'Insee⁶. L'étude porte donc sur les salaires réels, et tous les montants sont exprimés en euros constants 2011.

1.2 Les indicateurs de la précarité de l'emploi

Nous mobilisons trois indicateurs de la précarité de l'emploi⁷, qui, du plus restrictif au plus large, ne visent à identifier que cet aspect particulier de la notion protéiforme de précarité (Barbier, 2005). Le premier correspond au chômage au sens du Bureau international du travail (BIT) à la date de l'enquête. Le second, au fait de se déclarer spontanément chômeur à l'enquête, ou un an avant. Le troisième, d'appartenir à ce dernier groupe ou d'occuper un emploi précaire à la date de l'enquête.

Le chômage au sens du BIT retenu ici suit l'ancienne interprétation, moins restrictive que celle adoptée par l'Insee en 2003 en vue de l'harmonisation entre pays de l'Union européenne (Tableau 3.1). En effet, seule cette interprétation permet de reconstruire une série homogène depuis 1975 (Goux et Chardon, 2003). Selon la définition du BIT, est chômeur (ou plus précisément « personne sans emploi à la recherche d'un emploi ») un individu qui n'a pas travaillé au cours de la semaine de référence, était disponible pour travailler dans un délai de deux semaines, et a entrepris des démarches de recherche d'emploi au cours du mois précédant la semaine de référence⁸.

6 « Série historique du pouvoir d'achat euro/franc », <http://www.bdm.insee.fr/bdm2/choixCriteres.action?codeGroupe=1391>

7 Concept à distinguer notamment de celui de précarité du travail, qui porte quant à elle sur le contenu plutôt que sur la forme de l'activité professionnelle (Paugam, 2000).

8 L'ancienne comme la nouvelle interprétations se fondent sur les critères du BIT. Néanmoins, la nouvelle interprétation est plus exigeante concernant les démarches considérées comme spécifiques de la recherche d'emploi : en particulier, la seule inscription dans une agence de recherche d'emploi ne suffit plus pour être considéré comme chômeur.

Cette définition étant relativement restrictive, et peu appropriée à l'identification d'un potentiel groupe social constitué des individus souffrant de la précarité, nous recourons à une deuxième approche, incluant le « halo » du chômage (Cézard, 1986) dans son sens le plus large. Jusqu'en 2003, les questionnaires des enquêtes débutaient par un « filtre initial » demandant aux individus de définir de manière libre leur statut d'emploi. Cette définition dite spontanée du chômage offre une approche sociologiquement plus significative d'un potentiel groupe social des chômeurs, comprenant notamment les chômeurs dits découragés (Nauze-Fichet et Gonzalez-Demichel, 2003). Malheureusement, suivant les recommandations d'Eurostat, l'introduction d'une nouvelle définition plus stricte du chômage en 2003 s'est accompagnée de la suppression de cette question initiale (Goux et Chardon, 2003, p. 81), dont on pouvait effectivement craindre qu'elle n'incite les interrogés à mettre en cohérence leurs déclarations concernant leurs démarches de recherche d'emploi avec le statut qu'ils venaient de retenir. Elle est cependant toujours posée, à la fin du questionnaire, pour chacun des mois de l'année écoulée. C'est cette information que nous retenons, tout en reconnaissant que les questions très précises concernant les démarches de recherche d'emploi qui sont posées plus tôt (mais pas immédiatement avant la question concernée) peuvent avoir un effet restrictif sur cette déclaration, la rendant moins spontanée qu'auparavant. Les résultats présentés plus bas ne laissent cependant pas apparaître de rupture notable dans les séries.

Tableau 3.1 : Le codage des différentes définitions du chômage dans les enquêtes Emploi

Série	Définition du chômage, variable et modalité
1975-1981	BIT : UTA 2 Spontané : FI 2 Spontané un an auparavant : FIP 2
1982-1989	BIT : ACTBIT 2 et 3 Spontané : FI 2 et 3 Spontané un an auparavant : FIP 2 et 3
1990-2002	BIT : ACT 2 Spontané : FI 2 Spontané un an auparavant : FIP 2
2003-2011	BIT : ACTANC 2 Spontané* : SP00 4 jusqu'en 2007, SP00 3 ensuite Spontané un an auparavant : EOCCUA 2

* En 2003, la question portant sur le chômage « spontané » a été modifiée (voir le corps du texte).

L'ensemble des individus au chômage à un instant donné ne constituent pourtant pas un groupe stable. Les couples dans lesquels les deux conjoints se trouvent au chômage la même semaine de référence ne représentent qu'une partie relativement limitée de ceux dans lesquels les conjoints sont affectés par la précarité de l'emploi. Nous élargissons donc cette conception du halo du chômage en incluant les individus déclarant avoir été chômeurs un an auparavant, suivant la définition spontanée déjà décrite. S'il aurait été intéressant de retenir une période encore plus longue, cette information n'est pas disponible dans les enquêtes Emploi.

Enfin, de manière à saisir encore plus largement les situations de précarité de l'emploi, nous recourons à une troisième variable, qui – signe des temps – ne peut être construite à partir des enquêtes que depuis 1982 (voir l'annexe de Michel Villac dans Schnapper, 1989). Sont inclus dans son champ, en plus des chômeurs au sens de la deuxième définition, les contrats d'intérim, les contrats à durée déterminée (y compris le statut de contractuel dans la fonction publique), les emplois saisonniers, les contrats aidés, les vacations, les stages (à l'exception des fonctionnaires-stagiaires), ainsi que les contrats d'apprentissage (Tableau 3.2). Soulignons que, malgré l'extension importante du champ couvert par cette troisième variable, la précarité n'est définie ici que pour les individus qui se déclarent comme actifs, et

ne tient compte que du statut d'emploi. Or, la précarité de l'emploi peut aussi tenir aux risques de licenciement des individus en contrat à durée indéterminée dans des secteurs ou des professions instables ou en déclin. En outre, la précarité des emplois féminins s'exprime tout autant par l'inactivité et le temps partiel (« sous-emploi » dans les termes de Maruani, 2002), situations dont on mesure la difficulté d'appréhension dans un contexte d'expansion massive de l'activité des femmes : il n'est pas possible de considérer que l'ensemble des inactives de 1975 étaient des précaires⁹. Le temps partiel lui-même peut être choisi (dans la fonction publique notamment) autant que subi, bien qu'il soit dans les deux cas l'illustration de l'ordre de genre. Malgré ses limites, la catégorie ainsi définie semble plus à même que les précédentes de recouvrir un groupe souffrant durablement de l'instabilité de l'emploi, et potentiellement doté d'une certaine cohérence.

Tableau 3.2 : Le codage des emplois précaires dans les enquêtes Emploi

Série	Variable, modalité et intitulé
1982-1989	STATUT 3 (stagiaire en formation, rémunéré), 21 (intérimaire), 22 (apprenti sous contrat), 23 (stagiaire), 24 (saisonnier), 25 (autre CDD), 32 (contractuel avec limitation de durée), 33 (contractuel sans limitation de durée), 34 (auxiliaire), 35 (vacataire), 36 (autre non titulaire) TIT 2 (contractuel sans limitation de durée), 3 (contractuel pour une durée limitée), 4 (auxiliaire), 5 (vacataire), 6 (autre cas)
1990-2002	STATUT 21 (intérimaires hors État, collectivités locales), 22 (apprentis hors État, collectivités locales), 23 (CDD hors État, collectivités locales), 30 (stagiaires et contrats aidés) TIT 3 (contractuel sans limitation de durée), 4 (contractuel pour une durée limitée, auxiliaire, vacataire, pigiste)
2003-2011	STATUT 21 (intérimaires), 22 (apprentis), 33 (CDD hors État et collectivités locales, hors contrats aidés), 34 (stagiaires et contrats aidés, hors État et collectivités locales), 43 (CDD État et collectivités locales, hors contrats aidés), 44 (stagiaires et contrats aidés État et collectivités locales) TITC 3 (contractuel [<i>par opposition à titulaire ou fonctionnaire-stagiaire</i>])

L'échantillon est constitué de 314 000 couples depuis 1975, de 262 000 couples depuis 1982, et de 197 000 couples depuis 1990. Ces variables étant centrales pour l'enquête Emploi, et du fait des redressements portant sur le salaire, les valeurs manquantes sont quasiment inexistantes.

⁹ Si l'on considère l'inactivité comme un statut précaire, on mesure une diminution au cours du temps de la proportion de couples dans lesquels les deux conjoints correspondent à cette définition.

2 La diminution des inégalités de salaire entre couples et ses composantes

2.1 Les conséquences de l'homogamie et de la hausse de l'activité féminine sur les inégalités

La question de l'association entre les salaires des conjoints et de ses effets sur les inégalités a été soulevée par des travaux portant principalement sur les États-Unis : il s'agissait au départ d'évaluer l'effet de l'augmentation de l'activité professionnelle des femmes sur les inégalités entre ménages. En effet, comme on l'a déjà vu, le contexte américain diffère assez nettement de celui de la France puisqu'on y observe une forte augmentation de l'homogamie d'éducation, qui a fréquemment été perçue comme un risque pour la cohésion de la société américaine (Breen et Salazar, 2011). À rebours de ces anticipations, les travaux concordent pour conclure que la croissance de l'activité des femmes a au contraire eu tendance à limiter l'augmentation des inégalités aux États-Unis (Cancian et Reed, 1998, 1999 ; Reed et Cancian, 2001 ; Devereux, 2004 ; Pencavel, 2006 ; Western, Bloome et Percheski, 2008 ; Greenwood *et al.*, 2014 ; Hryshko, Juhn et McCue, 2014) ainsi que dans tous les pays de l'OCDE, et particulièrement en France (OCDE, 2011, p. 226).

En revanche, le rôle propre de l'évolution de l'homogamie est plus discuté. Si certains travaux ont avancé que l'augmentation de l'homogamie d'éducation explique une part importante de la croissance des inégalités entre ménages aux États-Unis (Greenwood *et al.*, 2014), d'autres ont affirmé que ces résultats ne peuvent être répliqués à partir des mêmes données (Hryshko, Juhn et McCue, 2014, p. 15, note 18 ; Harmenberg, 2014). La plupart des travaux concluent au contraire que l'homogamie d'éducation n'entretient qu'un lien relativement lâche, bien que réel, avec l'association entre les revenus des conjoints. Ainsi, contrairement à une opinion courante, l'évolution de l'homogamie d'éducation n'a que des effets négligeables – voire négatifs – sur l'augmentation des inégalités économiques (Worner, 2006 pour l'Australie ; Western, Bloome et Percheski, 2008 ; Breen et Salazar, 2010, 2011 pour les États-Unis et le Royaume-Uni ; Breen et Andersen, 2012 pour le Danemark ; Eika, Mogstad et Zafar, 2014 pour les États-Unis et la Norvège).

Il importe donc tout d'abord de s'attacher à identifier un éventuel effet de l'association entre salaires des conjoints sur les inégalités, avant de se poser la question de celui de l'homogamie d'éducation, qui n'est que l'un des déterminants de la première. Les travaux portant sur cette question observent soit que l'évolution de l'association entre revenus des conjoints explique 15 à 30 % de l'augmentation totale des inégalités aux États-Unis selon les périodes et les méthodes retenues (Karoly et Burtless, 1995 ; Burtless, 1999 ; Cancian et Reed, 1999 ; Hyslop, 2001 ; Schwartz, 2010), soit qu'elle y prend une part encore plus modeste (Cancian et Reed, 1998 ; Hryshko, Juhn et McCue, 2014 pour les États-Unis ; OCDE, 2011, p. 226 pour les pays membres de l'organisation ; Funes Leal, 2015 pour l'Argentine). En outre, aux États-Unis, ce facteur aurait principalement contribué à la hausse des inégalités dans les années 1980, mais de manière assez négligeable depuis (Larrimore, 2014).

Étant donné l'affaiblissement de l'homogamie selon plusieurs dimensions de la stratification sociale que nous avons déjà mis en évidence, il semble donc peu probable que l'évolution de ce facteur ait pu jouer un rôle significatif dans le sens d'une augmentation des inégalités entre ménages en France. Si la question de savoir si une baisse de l'association entre salaires des conjoints a pu prendre place reste à élucider, il paraît d'ores et déjà clair que l'effet égalisateur de la croissance de l'activité féminine l'emportera.

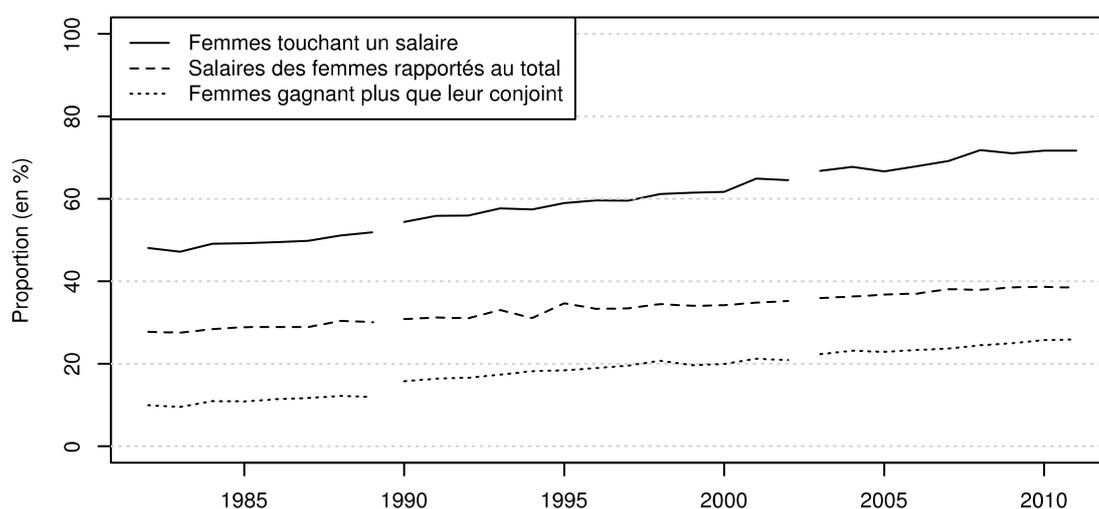
C'est bien ce dernier phénomène qui constitue, dans le cas de la France comme dans celui d'autres pays occidentaux, l'évolution majeure sur la période considérée. Entre 1982 et 2011, la part des femmes en couple cohabitant de 30 à 59 ans (hors indépendants) qui déclarent toucher un salaire est passée de 47 % à 72 % (Figure 3.1). Ces chiffres ne révèlent pas une autre évolution importante, invisible dans nos données transversales : la plus grande continuité temporelle des carrières féminines (Meron et Maruani, 2012). Plus encore, la part des salaires des femmes dans le total des salaires des couples a progressé presque au même rythme, passant de 28 % à 41 %. Enfin, la proportion des femmes gagnant plus que leur conjoint a plus que doublé, passant de seulement 10 % à 26 % (ce dernier chiffre étant similaire à celui obtenu par Morin, 2014).

Ces deux dernières évolutions s'expliquent principalement par l'augmentation de l'activité des femmes, mais aussi dans une certaine mesure par une augmentation – modérée mais réelle – des salaires des femmes par rapport à ceux des hommes. À titre d'illustration, l'écart entre salaires médians des hommes et des femmes en emploi, qui était de 27 % au

début de la période, a diminué jusqu'en 2004 puis s'est stabilisé aux alentours de 24 %. Ce mouvement assez faible a été plus marqué aux extrêmes de la distribution : on passe d'un écart de 66 % à un écart de 57 % en 2011 pour le premier décile, ainsi que de 40 % à 32 % pour le dernier décile, et aucune stabilisation n'est à relever.

L'évolution au cours des trente dernières années ne laisse pas entrevoir d'essoufflement de cette dynamique : les courbes sont presque parfaitement linéaires. Ainsi, en tenant compte de la rupture de série introduite en 1990¹⁰, on peut estimer que si le rythme d'augmentation très stable observé depuis 1982 se maintient, les salaires des femmes représenteront la moitié du total entre 2050 et 2055. De même, et de manière tout à fait cohérente avec le résultat précédent, la moitié des femmes gagneront plus que leur conjoint aux alentours de 2060. Ces estimations, si elles n'ont aucune prétention à la prédiction et peuvent paraître bien lointaines, indiquent que le rythme maintenu au cours des trente dernières années est assez soutenu.

Figure 3.1 : La croissance des salaires des femmes depuis 1982 à travers trois indicateurs



Source : enquêtes Emploi (Insee)

Champ : hommes et femmes vivant en couple cohabitant dans lequel l'un des conjoints a entre 30 et 59 ans et aucun n'est indépendant

Les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul.

¹⁰ À l'aide d'une variable indicatrice identifiant les années antérieures à 1990 dans un modèle de régression linéaire.

Cette évolution frappante n'empêche évidemment pas la persistance d'inégalités hommes femmes importantes (Maruani, 2000 ; Meron, 2008 ; Albouy, Djider et Mainguéné, 2012) : plus grande précarité des emplois féminins (notamment temps partiels), salaires inférieurs du fait de la structure de ces emplois, mais aussi à poste égal (*cf. supra*), ségrégation entre secteurs et professions¹¹. Elle semble cependant contredire les prédictions concernant l'essoufflement de la progression de l'emploi féminin, qui serait en réalité arrêtée en équivalent temps plein du fait du développement du temps partiel (Buffeteau et Afsa, 2006) : comme produit d'un salaire horaire et d'un temps de travail, le salaire réellement touché que nous retenons ici tient compte de l'évolution de ce facteur.

La réalité de l'évolution ne fait donc aucun doute. Nous allons maintenant chercher à évaluer si, comme dans d'autres pays, l'augmentation de plus de dix points de pourcentage de la part des salaires des femmes dans le total des salaires des couples a eu pour effet de réduire les inégalités entre couples, avant de chercher à mieux comprendre les mécanismes à l'œuvre.

2.2 Les trois composantes de l'inégalité entre salaires des couples

Afin de répondre à cette question, nous procédons dans cette section à une première décomposition de l'inégalité entre salaires des couples, et de son évolution temporelle, en trois sources distinctes. Nous mesurons pour cela le niveau d'ensemble des inégalités à l'aide du coefficient de variation, qui présente une définition aisément interprétable en termes de dispersion par rapport à la moyenne. Il est défini comme le rapport entre l'écart-type σ et la moyenne μ d'une distribution, soit :

$$CV = \frac{\sigma}{\mu}$$

De nombreux indicateurs des inégalités ont été utilisés dans la littérature, la plupart étant des cas particuliers de l'indice d'entropie généralisée, satisfaisant le même ensemble de propriétés attendu d'un tel outil (Bourguignon, 1979 ; Shorrocks, 1980) : c'est le cas de l'indice de Gini, de l'indice de Theil, et du coefficient de variation. Nous avons retenu ce dernier comme mesure du niveau global des inégalités par cohérence avec une grande partie de la littérature portant spécifiquement sur le lien entre salaires des conjoints, utilisation qui

11 En revanche, le surcroît de chômage des femmes s'est récemment résorbé à la suite de l'augmentation de celui des hommes due à la crise économique commencée en 2008 (Cabannes, 2014).

s'explique elle-même par la souplesse et la simplicité de sa décomposition. Des mesures permettant d'identifier des évolutions non uniformes (variables selon la partie de la distribution considérée) seront utilisées dans la section suivante.

L'évolution de l'inégalité entre les salaires totaux des couples (en attribuant un salaire nul aux chômeurs et inactifs) est présentée dans la figure 3.2 (courbe épaisse) : on observe une baisse limitée mais relativement régulière du niveau global des inégalités¹², le coefficient de variation passant de 0,50 à 0,44 entre 1990 et 2011, soit une diminution de 14 %. Cette évolution est cohérente avec celle observée par les travaux existants, qui portent sur l'ensemble des sources de revenus et tiennent compte des différences de taille des ménages (Pujol et Tomasini, 2009 ; OCDE, 2008, p. 29, 2011, p. 215). La décomposition du coefficient de variation va nous permettre d'identifier les sources de cette diminution.

En effet, le carré du coefficient de variation peut être écrit comme la somme de trois termes correspondant chacun à une source de revenu bien identifiée¹³ :

$$CV^2 = (1 - S_f)^2 CV_h^2 + S_f^2 CV_f^2 + 2\rho_{hf} S_f (1 - S_f) CV_h CV_f$$

S_f mesure la part des salaires des femmes dans le total des salaires des couples, CV_h et CV_f sont respectivement les coefficients de variation des salaires des hommes et des femmes, et ρ_{hf} correspond à la corrélation entre les salaires des conjoints au sein des couples. L'inégalité entre couples est donc d'autant plus prononcée que l'inégalité entre les individus de chaque sexe est élevée, que le sexe qui connaît l'inégalité la plus marquée contribue fortement au total des salaires des couples, et que la corrélation entre salaires des conjoints est forte.

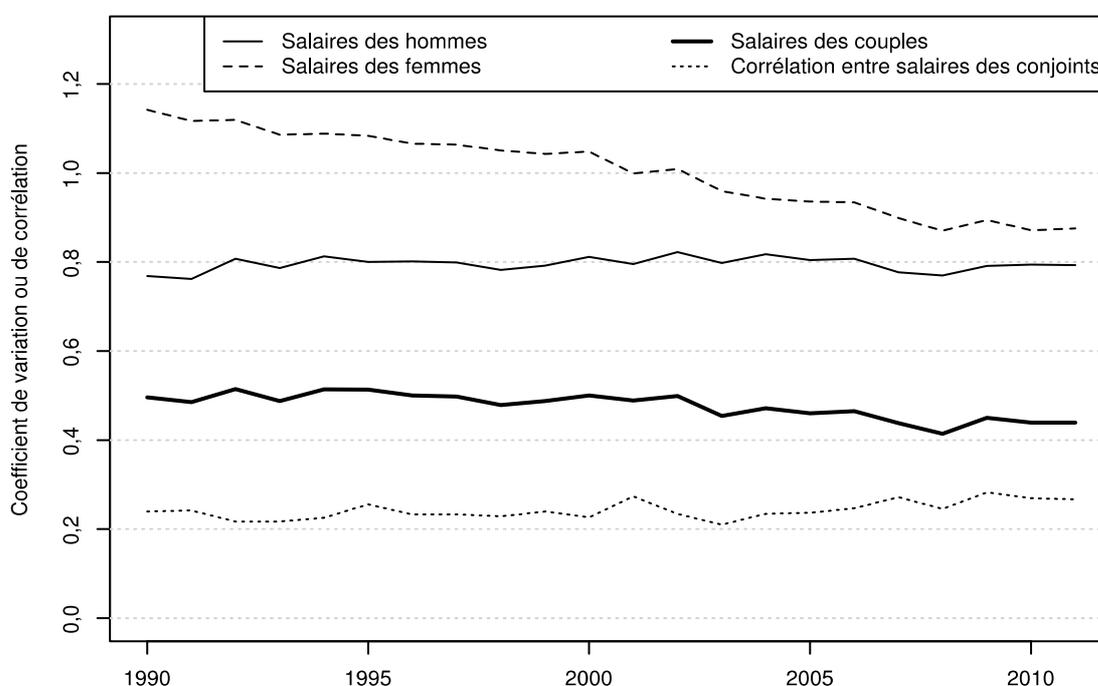
D'après cette équation, la contribution de chaque sexe aux inégalités entre couples peut être évaluée par rapport à (au moins) trois situations de référence dites contrefactuelles, qui permettent d'identifier trois sources différentes de l'évolution des inégalités (Cancian et Reed, 1998, p. 74). Première situation : si seuls les individus d'un sexe contribuaient au total des salaires des couples, les membres de l'autre sexe étant tous inactifs ou chômeurs, le niveau des inégalités entre couples serait respectivement de CV_h ou de CV_f selon que les

12 On peut cependant relever un léger rebond en 2008 suite au déclenchement de la crise économique, puis une stabilisation.

13 Cette décomposition est obtenue par l'application d'une identité remarquable à la formule de la covariance.

actifs occupés sont les hommes ou les femmes¹⁴. Deuxième situation : si les salaires des individus d'un sexe donné étaient tous égaux, et que par conséquent l'inégalité trouvait exclusivement son origine chez l'autre sexe, alors le niveau d'inégalité entre les couples serait égal à $(1 - S_f)CV_h$ ou S_fCV_f , selon que l'inégalité proviendrait des hommes ou des femmes¹⁵. Troisième situation : si cette hypothèse était elle aussi relâchée, mais qu'il n'existait aucune corrélation entre les salaires des conjoints, le coefficient de variation serait égal à la racine carrée de la somme des termes $(1 - S_f)^2CV_h^2$ et $S_f^2CV_f^2$.

Figure 3.2 : Évolution des inégalités de salaire et de la corrélation entre salaires des conjoints depuis 1990



Source : enquêtes Emploi (Insee)

Champ : hommes et femmes vivant en couple cohabitant dans lequel l'un des conjoints a entre 30 et 59 ans et aucun n'est indépendant

Les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul.

Les séries « salaires des hommes » et « salaires des femmes » représentent le coefficient de variation correspondant aux salaires des individus vivant en couple du sexe concerné. La série « salaires des couples » représente le coefficient de variation de la somme des salaires des conjoints. La série « corrélation entre conjoints » représente la corrélation entre salaires de l'homme et de la femme au sein des couples.

14 En effet, on peut vérifier que l'équation ci-dessus se simplifie radicalement lorsque S_f est égal à 0 ou à 1.

15 Cette situation correspond au cas dans lequel l'un des termes CV_h et CV_f est nul.

Ainsi, l'association entre les salaires des conjoints joue potentiellement un rôle important dans la détermination des inégalités entre couples. Dans une situation hypothétique, irréaliste en 1990 mais de plus en plus crédible avec le temps, où les inégalités de salaire seraient les mêmes chez les hommes et chez les femmes et où les deux sexes contribueraient à parts égales au volume total des salaires, passer d'une absence de corrélation à une corrélation parfaite entre salaires des conjoints équivaldrait exactement à doubler le niveau des inégalités entre couples. Cependant, une évolution aussi radicale de la corrélation est assez improbable : comme on l'a souligné, l'augmentation du taux d'emploi des femmes a toutes les chances de constituer le facteur le plus déterminant de l'évolution des inégalités.

De fait, la légère baisse de l'inégalité d'ensemble s'explique par le fait que les différentes composantes de l'équation sont soit restées relativement stables, soit ont évolué vers moins d'inégalités (Figure 3.2). Tout d'abord, l'inégalité entre les salaires des hommes (CV_h) a peu évolué, fluctuant autour d'un coefficient de variation de 0,8¹⁶. En revanche, les inégalités entre les salaires des femmes (CV_f) ont très nettement diminué, avec un coefficient de variation passant de 1,15 à 0,88 : elles se sont très nettement rapprochées du niveau qui prévaut chez les hommes. Ce mouvement peut s'expliquer par le développement de l'activité des femmes, qui a fait passer le salaire d'une partie importante de l'échantillon d'une valeur nulle (inactivité) à un niveau qui a toutes les chances d'être plus proche de la moyenne. Cette diminution des inégalités entre femmes s'est conjuguée avec la nette augmentation de la part des salaires des femmes dans les salaires des couples (S_f , cf. Figure 3.1, p. 195) pour entraîner la baisse des inégalités entre couples déjà commentée.

Si cette évolution est ainsi une conséquence mécanique de l'augmentation du taux d'emploi féminin, il est particulièrement intéressant de remarquer qu'elle ne s'est pas accompagnée d'une nette augmentation du troisième terme de l'équation – la corrélation entre salaires des conjoints –, qui aurait pu annuler pour partie ou complètement son influence négative sur les inégalités entre couples. De fait, cette corrélation (actifs et inactifs confondus) a peu varié depuis 1990, se situant légèrement en dessous de 0,25. Elle semble

16 Des analyses recourant à d'autres indicateurs observent une légère baisse des inégalités de salaire, dues principalement à l'augmentation relative des bas salaires (Verdugo, Fraisse et Horny, 2012). L'étude des niveaux de vie en tenant compte de l'ensemble des revenus fournit des résultats similaires (Pujol et Tomasini, 2009). Soulignons que ces études, comme la nôtre, ne sont pas en mesure d'étudier les inégalités au sommet de la distribution.

avoir connu une très légère augmentation depuis 2003 – dont on ne peut cependant exclure qu'elle soit due au rodage du nouveau dispositif de l'enquête Emploi en continu. Ce niveau peut paraître relativement faible ; il est cependant nettement plus élevé que celui rapporté par plusieurs auteurs concernant les États-Unis, pays dans lequel la corrélation était légèrement négative¹⁷ avant 1980, et s'établit autour de 0,1 dans les années récentes (Schwartz, 2010, p. 1540 ; Cancian et Reed, 1998, p. 76 ; Reed et Cancian, 2012, p. 10). On peut y voir un signe du modèle français d'activité des femmes, dans lequel l'emploi à temps plein occupe une place plus importante que dans d'autres pays (Maruani, 2000 ; Meron et Maruani, 2012).

Ce scénario peut paraître assez inattendu – il s'oppose par exemple à une conjecture de Thomas Piketty (2003, p. 1018, note 16) : en effet, la corrélation entre les salaires des conjoints semble *a priori* toujours devoir être plus forte parmi les couples dans lesquels le salaire de la femme n'est pas nul et peut par conséquent varier selon le statut social du couple. L'existence d'une forte proportion d'inactives paraît donc nécessairement tirer la corrélation entre salaires des conjoints vers le bas. À l'appui de cette intuition, la corrélation entre salaires des conjoints au sein des couples dans lesquels la femme est active s'établit autour de 0,3, soit légèrement plus que celle observée dans l'ensemble des couples¹⁸ ; cet écart a aussi été noté aux États-Unis (Schwartz, 2010), avec ici encore une corrélation moins forte qu'en France.

Mais l'effet de l'augmentation de la proportion de couples bi-actifs sur la corrélation entre salaires des conjoints parmi l'ensemble des couples dépend en grande partie du lien entre inactivité de la femme et niveau de salaire de son conjoint. Il ne fait aucun doute que l'augmentation de l'activité des femmes aurait dû se traduire par un renforcement de la corrélation entre salaires des conjoints si la proportion d'actives avait le plus augmenté parmi les conjointes des hommes percevant les salaires les plus élevés. Ainsi, aux États-Unis, l'augmentation de la corrélation entre salaires des conjoints s'explique en grande partie par le fait que la relation négative entre activité de la femme et salaire de son conjoint a progressivement disparu, laissant place à une courbe en U inversé suivant laquelle les classes moyennes présentent les plus forts taux d'emploi féminin (Schwartz, 2010, p. 1541).

17 Du fait notamment du taux d'inactivité élevé des conjointes des hommes appartenant aux derniers déciles (*cf. infra*).

18 On peut noter une relative convergence entre les deux groupes, qui s'explique par la proportion décroissante d'inactives.

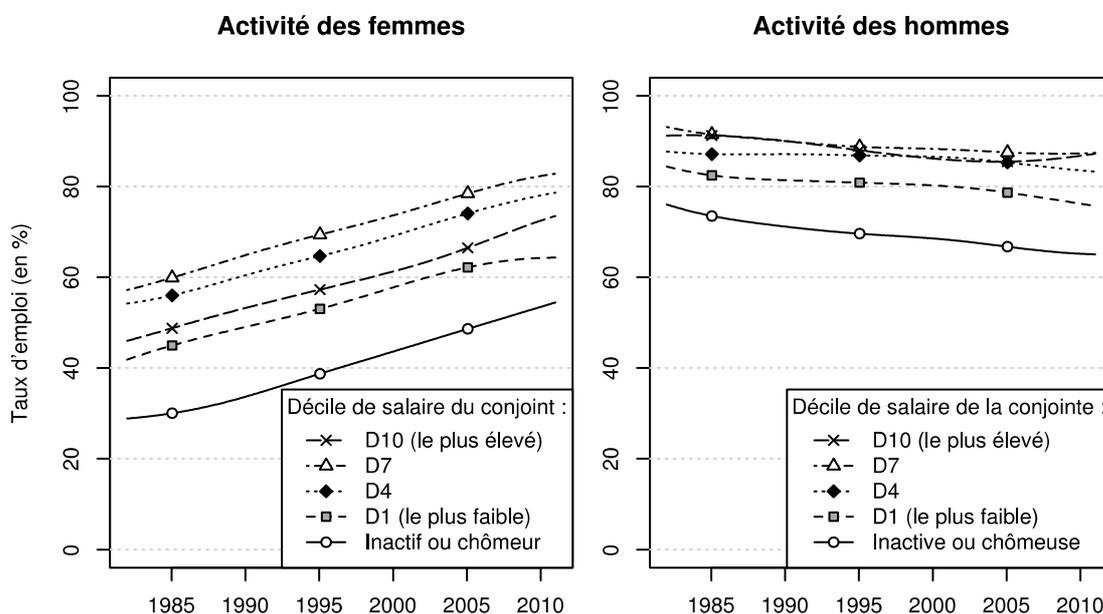
En revanche, si en France l'activité féminine s'est développée de manière homogène entre les groupes sociaux, alors cette évolution a pu n'avoir qu'un effet limité sur l'association entre salaires des conjoints.

2.3 La distribution sociale de l'inactivité

On remarque en effet que le cas français est relativement différent du cas américain (Figure 3.3, graphique de gauche). Dès 1982, le taux d'emploi des femmes le plus élevé (à près de 60 %) se rencontrait parmi les conjointes des hommes appartenant au septième décile de salaire. Son niveau était presque aussi faible pour le dernier décile (11 points de pourcentage en dessous du maximum) que pour le premier (14 points en dessous). Néanmoins, les conjointes des hommes ne touchant aucun salaire se distinguaient nettement par un taux d'emploi inférieur au maximum de près de 30 points. Aux États-Unis à la même période, seuls les déciles supérieurs se distinguaient par un taux d'emploi féminin clairement plus faible que les autres, les déciles inférieurs les ayant progressivement rejoints seulement après cette date. La situation française semble donc ici avoir anticipé l'évolution observée aux États-Unis : en France, l'activité des femmes avait déjà pour effet d'augmenter les salaires des couples du milieu de la distribution plus que ceux des couples situés à ses extrêmes. Ainsi, l'augmentation du taux d'emploi des femmes ayant affecté de manière homogène tous les déciles, elle n'a pas eu pour conséquence un renforcement des inégalités entre couples. Malheureusement, nos données ne nous permettent pas de savoir si le mouvement relevé aux États-Unis s'est produit en France avant les années 1980, ou si l'échelle que nous observons était déjà en place de longue date.

Il n'est pas non plus inintéressant d'étudier les variations de l'activité des hommes selon le salaire de leur conjointe (Figure 3.3, graphique de droite). Outre qu'une légère baisse d'ensemble du taux d'emploi contraste avec la hausse relevée chez les femmes, le tableau est assez différent. D'une part, le taux d'emploi des hommes croît avec le salaire de leur conjointe jusqu'à la médiane, et se stabilise ensuite. D'autre part, les écarts sont ici bien plus compressés, avec un maximum autour de 20 points de pourcentage.

Figure 3.3 : Évolution du taux d'emploi des hommes et des femmes selon le salaire de leur conjoint



Source : enquêtes Emploi (Insee)

Champ : hommes et femmes vivant en couple cohabitant dans lequel l'un des conjoints a entre 30 et 59 ans et aucun n'est indépendant

Les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul.

Les courbes sont lissées à l'aide d'une régression locale de degré 1 (LOESS). Cette technique de lissage non paramétrique consiste à estimer pour chaque année la position verticale du point à partir d'une droite de régression calculée en se fondant sur l'ensemble de la courbe observée, en attribuant des poids décroissants aux points à mesure que l'on s'éloigne de l'année considérée (les poids suivent ici un noyau gaussien).

Chez les deux sexes, le groupe dont les conjoints ne touchent aucun revenu se détache toujours assez nettement des autres. L'écart entre le taux d'emploi de ce groupe et le taux maximum tend à se renforcer légèrement au cours du temps chez les hommes (passant de 15 à 22 points) ; le même mouvement semble se produire dans les années récentes, chez les deux sexes, en ce qui concerne le premier décile. Dans l'ensemble cependant, les évolutions sont essentiellement parallèles.

Ainsi, l'évolution du taux d'emploi des femmes et des hommes n'a pas représenté en France une source importante d'évolution de la corrélation entre salaires des conjoints comme elle l'a été aux États-Unis, ce qui explique la grande stabilité de cette dernière. C'est l'homogénéité à travers les groupes sociaux de l'accroissement du taux d'emploi des femmes

qui a permis à cette évolution de jouer son rôle égalisateur sans se voir contrariée par une augmentation de la corrélation entre salaires des conjoints. Dans la section suivante, une décomposition plus détaillée des inégalités nous permettra de confirmer cette interprétation, ainsi que d'évaluer si elle a affecté de manière uniforme l'ensemble de la distribution des salaires des couples.

3 Les facteurs de l'évolution des inégalités : scénarios contrefactuels et évolutions localisées

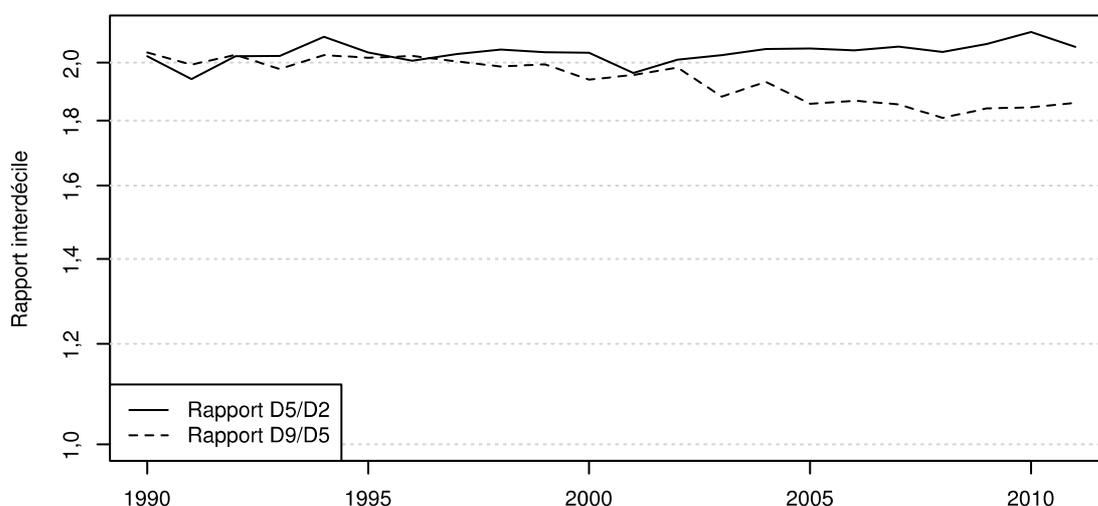
La décomposition du coefficient de variation utilisée dans la section précédente présente l'avantage d'une grande simplicité de mise en œuvre. En revanche, elle ne permet pas de mettre en évidence des effets potentiellement non uniformes selon les segments de la distribution des salaires, ni d'évaluer l'évolution des inégalités qui aurait été observée dans d'autres situations contrefactuelles que celles déjà évoquées. Dans cette section, nous nous inspirons de la démarche élaborée par Christine Schwartz (2010), qui consiste à modéliser la distribution jointe des revenus des deux conjoints à l'aide de modèles log-linéaires. Ceux-ci permettent, en imposant des contraintes aux paramètres, de simuler plusieurs situations contrefactuelles, et ainsi d'évaluer la contribution de chacune des différentes tendances à l'évolution des inégalités.

Beaucoup plus souple que la précédente, cette décomposition ne se limite plus nécessairement au coefficient de variation. Afin d'étudier des évolutions potentiellement différentes selon la partie de la distribution des salaires des couples considérée, nous recourons aussi à deux rapports interdéciles : le rapport entre le décile supérieur et la médiane ($D9/D5$), et celui entre la médiane et le deuxième décile ($D5/D2$)¹⁹ ; le produit de ces deux rapports fournit le rapport entre le dernier et le deuxième déciles ($D9/D2$). L'intérêt de cette approche plus détaillée apparaît dans la figure 3.4 : on remarque que les inégalités dans la moitié supérieure de la distribution ($D9/D5$) ont diminué d'environ 10 %, alors que celles dans sa moitié inférieure sont restées très stables, et pourraient même avoir légèrement augmenté depuis 2001, notamment entre 2008 et 2010. La diminution du niveau global

19 C'est souvent le premier décile (D1) qui est retenu plutôt que le deuxième. Ce choix n'est pas possible ici : dans la mesure où un peu plus de 10 % des couples ne touchent aucun salaire, le rapport $D5/D1$ n'est pas défini. On voit donc que le rapport $D5/D1$, lorsqu'il est calculé en tenant compte des revenus de transfert, mesure en fait le fonctionnement du système socio-fiscal, par le biais de la redistribution vers les ménages les plus pauvres, plus que celui du marché du travail.

d'inégalité entre couples identifiée dans la partie précédente à l'aide du coefficient de variation semble donc provenir uniquement d'une compression par le haut des inégalités entre couples. Nous chercherons dans ce qui suit à comprendre ce phénomène en identifiant les facteurs qui en sont à l'origine.

Figure 3.4 : Évolution de deux rapports interdéciles entre salaires des couples depuis 1990



Source : enquêtes Emploi (Insee)

Champ : hommes et femmes vivant en couple cohabitant dans lequel l'un des conjoints a entre 30 et 59 ans et aucun n'est indépendant

Les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul.

3.1 La méthode de décomposition par la modélisation log-linéaire

La méthode de décomposition proposée par C. Schwartz (2010) consiste à découper la distribution des salaires de chaque sexe observée chaque année en quantiles relativement fins, plus un groupe pour les inactifs et les chômeurs, et de construire une table d'homogamie à trois dimensions croisant chaque année les quantiles d'appartenance de chacun des conjoints. Le coefficient de variation et les rapports interdéciles peuvent être recalculés à partir de cette table, en attribuant à chaque individu le salaire médian de son quantile d'appartenance. Un découpage détaillé permet de reproduire de manière assez fidèle les valeurs des indices mesurées à partir des données brutes.

Comme cet auteur, nous avons ici choisi de découper la distribution en 20 quantiles représentant chacun 5 % des individus de chaque sexe ayant touché un salaire. Ce niveau de précision va en effet au-delà de ce qui est requis pour calculer un rapport interdéciles, et permet au coefficient de variation de saisir les inégalités au sommet de manière relativement correcte au regard de la taille de l'échantillon dont nous disposons. Nous avons vérifié que les résultats obtenus en utilisant la table réellement observée (incluant donc l'approximation liée au découpage) étaient presque exactement identiques aux résultats bruts présentés plus haut, ainsi qu'à ceux obtenus avec un découpage plus fin en percentiles.

L'objectif de cette méthode est de modéliser, à partir de la table d'homogamie réellement observée, plusieurs tables, correspondant chacune à une situation contrefactuelle. Grâce à la flexibilité des modèles log-linéaires, il est ainsi possible d'élaborer des hypothèses aussi complexes que nécessaire. Plus précisément, il s'agit de partir d'un modèle imposant des hypothèses très fortes sur l'association entre salaires des conjoints et son évolution, et de les relâcher peu à peu par l'introduction de termes supplémentaires : chaque modèle est une extension du précédent. En comparant les valeurs des indices d'inégalité obtenus pour chaque situation, on obtient une estimation de la contribution de la levée de chacune des hypothèses aux inégalités et à leur évolution temporelle. Nous commençons par décrire les modèles et les situations auxquelles ils correspondent, avant de présenter les résultats de cette décomposition (on trouvera les équations des modèles dans l'encadré 3.1).

Le premier modèle suppose qu'il n'existe aucune association entre les salaires des conjoints, ni entre activité de l'un des conjoints et salaire de l'autre : seule la proportion d'actifs et d'inactifs est autorisée à varier chaque année. Ce modèle est celui d'indépendance conditionnelle (*cf.* chapitre précédent), qui ne contient de paramètres que pour reconstituer les proportions marginales, et leur permettre de varier au cours du temps. À partir des effectifs prédits dans chaque cellule, il est possible de calculer les inégalités entre couples qui seraient observées dans deux situations différentes. Dans la première, on suppose que les salaires médians de chaque quantile (reflétant la distribution des salaires parmi les actifs de chaque sexe pris séparément) sont restés les mêmes entre 1990 et 2011. Dans la seconde, on retient les salaires médians réellement observés en 2011, reflétant ainsi l'évolution des inégalités parmi les individus de chaque sexe. Ces deux premières situations nous permettent d'évaluer la contribution de l'augmentation de l'activité des femmes (et,

secondairement, la légère baisse de celle des hommes) à l'évolution des inégalités entre couples (première situation), ainsi que celle de l'évolution des inégalités parmi chaque sexe (comparaison de la seconde situation à la première).

La troisième situation suppose que la seule association entre salaires des conjoints vient du lien entre activité de la femme et salaire de son conjoint, tout en supposant que ce lien est resté stable au cours du temps (modèle 2). Elle permet de mesurer l'effet déjà relevé de la distribution socialement marquée de l'inactivité féminine sur les inégalités et leur évolution.

La quatrième situation étend la précédente en supposant qu'il existe une association entre salaires des conjoints (y compris l'inactivité), mais que celle-ci est restée stable au cours du temps (modèle 3). Contrairement à ce que l'on pourrait attendre, ce modèle ainsi que le précédent prédisent une évolution des inégalités au cours du temps différente du premier modèle. En effet, même si l'association elle-même reste stable, l'évolution des distributions marginales de la table (répartition des hommes et des femmes entre actifs et inactifs) déplace les effectifs entre des zones de la table (couples inactifs, couples composés d'un seul actif, couples bi-actifs) qui diffèrent par l'association prédite. Or, contrairement à l'*odds ratio*, le coefficient de variation et les rapports interdéciles ne sont pas des indicateurs indépendants des marges. Ainsi, la comparaison de l'inégalité observée sous ce modèle à celle observée sous le modèle précédent permet d'évaluer l'effet de l'existence d'une association entre salaires des conjoints sur l'évolution des inégalités, même en supposant que cette association est restée stable.

La cinquième situation introduit une première source d'évolution temporelle de l'association entre salaires des conjoints : celle du lien entre activité de la femme et salaire de l'homme (modèle 4). Ce modèle permet de mesurer la contribution spécifique de l'évolution de ce lien, sans encore tenir compte de l'effet propre de l'évolution de l'association entre salaires des conjoints, qui est encore supposée rester stable. L'évolution temporelle est supposée être linéaire. En effet, cette spécification apparaît comme la meilleure d'après les indicateurs BIC et AIC, ce qui est cohérent avec la linéarité presque parfaite des évolutions déjà commentées plus haut, et l'absence de rupture entre séries.

Enfin, la sixième et dernière situation correspond à la table réellement observée (modèle saturé). La différence par rapport à la situation précédente provient donc de l'évolution au

cours du temps de l'association entre salaires des conjoints parmi les couples dans lesquels la femme est active.

Pour tous les modèles, puisque nous voulons évaluer la contribution de chaque facteur à l'évolution globale des inégalités entre 1990 et 2011, nous ne comparons pas directement les valeurs des indices correspondant à ces années : en effet, ces dernières contiennent une part de variation liée à l'échantillonnage aléatoire ainsi qu'à la conjoncture économique. Afin de calculer la contribution moyenne sur la période, nous retenons la valeur des indices prédite pour les deux années de référence par une régression linéaire simple de l'indice sur l'année d'enquête, ce qui revient à lisser l'évolution temporelle pour en extraire la tendance linéaire.

Encadré 3.1 : Équations des modèles log-linéaires utilisés pour la décomposition

Le premier modèle, celui d'indépendance conditionnelle, ne tient compte que de l'évolution de la distribution marginale des hommes et des femmes au cours du temps. En notant m_{hft} les effectifs prédits par le modèle pour la cellule à l'intersection de la ligne h (quantile de l'homme), de la colonne f (quantile de la femme) et de la couche t (année d'enquête), dans une table de dimensions $H \times F \times T$, ce modèle s'écrit :

$$M1 : \log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT}$$

Étant donné que les populations des deux sexes sont découpées chaque année en quantiles, les paramètres marginaux sont à ce stade de peu d'effet (à l'approximation liée au découpage près), à l'exception de ceux reflétant les proportions marginales d'inactives et d'inactifs : c'est pourquoi ce modèle revient principalement à autoriser ces proportions à varier au cours du temps.

Le second modèle ajoute à la spécification du précédent l'association entre l'inactivité de la femme et quantile de salaire de l'homme, en supposant qu'elle est restée stable dans le temps. Il s'écrit :

$$M2 : M1 + \lambda_h^{HF_0} \mathbb{1}_{f=0}$$

avec $f = 0$ indiquant l'absence de salaire de la femme (inactive ou chômeuse), et $\mathbb{1}_{f=0}$ l'indicatrice prenant la valeur 1 lorsque la femme est inactive et 0 si elle est active (touche un salaire).

Le troisième modèle, celui de stabilité de l'association, inclut l'association complète mais stable entre salaires de l'homme et de la femme. Il intègre donc l'association entre activité et salaire du conjoint du modèle précédent (qui n'apparaît cependant plus comme un terme spécifique). Il s'écrit :

$$M3 : MI + \lambda_{hf}^{HF}$$

Bien que ces trois premiers modèles supposent que l'association est restée stable en termes d'*odds ratios*, les indices d'inégalité peuvent varier au cours du temps, puisqu'ils ne sont pas indépendants des marges.

Le quatrième modèle autorise en plus l'association entre l'inactivité de la femme et le quantile de salaire de l'homme à varier de manière linéaire selon les années. Il s'écrit :

$$M4 : M3 + (\lambda^{F_0} + t\lambda^{F_0T})\lambda_h^{HF_0T} \mathbb{1}_{f=0}$$

avec $f = 0$ indiquant l'absence de salaire de la femme.

3.2 Résultats des modèles : les contributions aux inégalités

Pour radicale qu'elle soit, l'hypothèse d'une absence d'association entre salaires des conjoints décrit déjà raisonnablement bien les données (Tableau 3.3) : seuls 18 % des couples sont mal classés par le modèle d'indépendance conditionnelle (modèle 1). Cette observation rejoint le constat déjà établi au chapitre précédent concernant d'autres dimensions de l'homogamie. Dans cette première situation contrefactuelle, selon laquelle les inégalités de salaire entre hommes et entre femmes n'auraient pas évolué, seule la proportion d'actifs étant autorisée à varier, les inégalités entre couples auraient décliné de manière importante autant au niveau global (coefficient de variation) que dans la partie supérieure de la distribution (D9/D5) entre 1990 et 2011 (Figure 3.5). En revanche, elles seraient restées

stables dans la partie inférieure de la distribution (D5/D2) – comme dans tous les scénarios (Figure 3.5) et dans les données réellement observées (Figure 3.4, p. 204).

Tableau 3.3 : Indicateurs d’ajustement des modèles utilisés pour la décomposition

	D. L.	Déviante	Δ (%)	BIC	AIC
M1 : Indépendance conditionnelle	8 800	41 455	17,98	-65 842	23 855
M2 : M1 + Association activité-salaire stable	8 780	30 977	12,65	-76 076	13 417
M3 : M2 + Association complète stable	8 400	10 722	7,83	-91 698	-6 078
M4 : M3 + Association activité-salaire variable	8 380	10 689	7,78	-91 487	-6 071

D. L. : degrés de liberté. Δ : indice de dissimilarité.

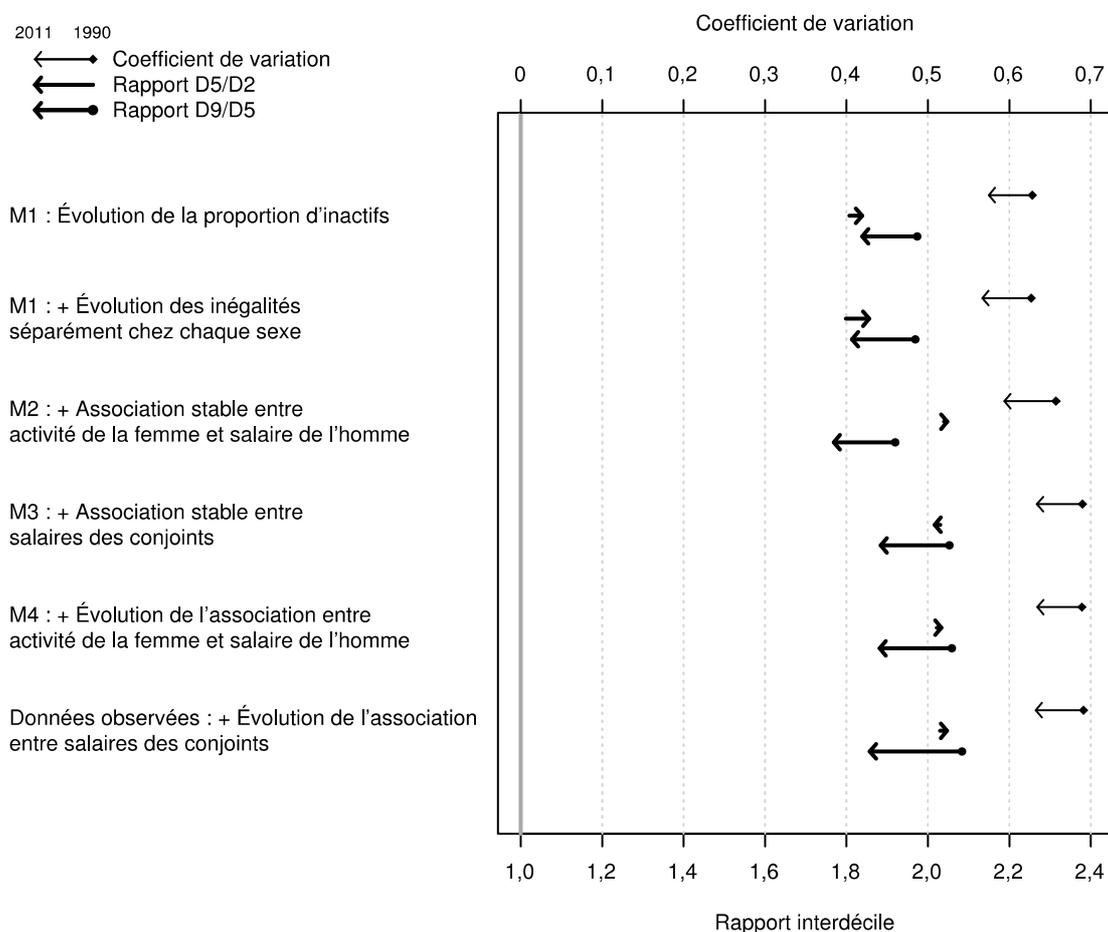
La baisse des deux premiers indicateurs est due à l’effet brut de la diminution de la proportion de femmes inactives : l’évolution de la seule activité des hommes n’aurait entraîné aucune variation des inégalités entre couples²⁰. On voit que ce facteur suffit à lui seul à expliquer la plus grande part des évolutions observées : leur ampleur est largement similaire dans les scénarios contrefactuels suivants. Ainsi, si les inégalités dans la partie inférieure de la distribution ont très peu évolué, c’est qu’elles sont très peu affectées par ce déterminant. On semble ici en présence d’un mécanisme analogue au phénomène classique de régression vers la moyenne (Verbakel, 2008, p. 132)²¹ : en effet, dans le scénario de mise en couple aléatoire retenu ici, il est d’autant moins probable que les deux conjoints touchent des salaires proches que ceux-ci sont éloignés du salaire moyen. La croissance du taux d’emploi des femmes entraîne donc une augmentation du salaire total des couples nettement plus forte (en termes relatifs) au niveau de la médiane qu’au niveau des quantiles extrêmes. En outre, étant donné que les variations sont bien plus amples au sommet qu’au bas de la distribution des salaires, cet effet s’exprime bien plus fortement dans le rapport D9/D5 que dans le rapport D5/D2 – et, ce, d’autant plus que les salaires des femmes sont plus compressés que ceux des hommes au sommet de la distribution. De fait, le rapport D5/D2

20 Ces deux situations contrefactuelles, omises ici pour limiter la complexité de la démarche, correspondent à des modèles autorisant la distribution marginale d’un seul sexe à varier au cours du temps.

21 La régression vers la moyenne désigne le phénomène statistique selon lequel, lors de la répétition d’un événement, les individus ayant obtenu les résultats les plus extrêmes la première fois tendront à se rapprocher de la moyenne la seconde. Ce principe s’applique par exemple à la taille des enfants par rapport à celle de leurs parents, ou encore aux résultats scolaires ou sportifs d’une épreuve sur l’autre. Par extension, chacun des conjoints peut être considéré comme une réalisation sur le marché du travail du statut social du couple.

n'est presque pas affecté par l'évolution de l'activité féminine, et augmente même très légèrement dans ce scénario. C'est le mérite de cette situation contrefactuelle que de souligner les effets contrastés de l'augmentation de l'activité féminine selon la partie de la distribution considérée.

Figure 3.5 : Évolution du coefficient de variation et des rapports D5/D2 et D9/D5 entre 1990 et 2011 dans les différentes situations contrefactuelles



Source : enquêtes Emploi (Insee)

Champ : hommes et femmes vivant en couple cohabitant dans lequel l'un des conjoints a entre 30 et 59 ans et aucun n'est indépendant

Les chômeurs et inactifs sont inclus dans le champ et se voient attribuer un salaire nul.

En recourant toujours au modèle d'indépendance conditionnelle (modèle 1), la deuxième situation contrefactuelle, qui autorise cette fois les inégalités de salaire à évoluer pour chaque sexe séparément au cours du temps, décrit une baisse des inégalités légèrement plus forte encore concernant les deux premiers indicateurs, mais une hausse des inégalités un peu plus marquée dans la partie inférieure de la distribution. Ce dernier mouvement, assez faible, s'explique par la baisse des salaires masculins les plus faibles par rapport à la médiane. Au contraire, le renforcement de la baisse observé pour les deux autres indicateurs, lui aussi peu marqué, s'explique par la diminution des inégalités chez les deux sexes, mais particulièrement chez les femmes²².

Mais ces deux scénarios ignorent un facteur important des inégalités entre couples : l'association inactivité de la femme et salaire de l'homme. Ce facteur est déterminant : le modèle 2, qui tient compte de cette association sans l'autoriser à varier au cours du temps, présente un indice de dissimilarité d'un peu moins de 13 % (Tableau 3.3), inférieur de cinq points à celui du modèle d'indépendance conditionnelle. Cette association présente des effets divergents sur le niveau des inégalités selon l'indicateur considéré. Du fait de la sur-représentation des inactives parmi les conjointes des hommes appartenant au dernier décile de salaire, elle entraîne une diminution du niveau des inégalités dans la partie supérieure de la distribution (D9/D5). Au contraire, les inégalités d'ensemble (coefficient de variation), et surtout celles dans la partie inférieure de la distribution (D5/D2), augmentent à cause de la corrélation positive entre activité de la femme et salaire de son conjoint qui existe dans le reste de la distribution (Figure 3.3, p. 202). En revanche, l'effet de cette association sur l'évolution des inégalités entre 1990 et 2011 est assez faible.

La quatrième situation contrefactuelle (modèle 3) tient compte de l'existence d'une association entre salaires des conjoints (notamment parmi les couples dans lesquels la femme occupe un emploi), mais suppose toujours que cette association est restée stable dans le temps. Ce modèle améliore encore nettement la description des données (Tableau 3.3), en diminuant de près de 5 points l'indice de dissimilarité, signe que l'association entre salaires des conjoints n'est pas un phénomène négligeable. Par rapport au précédent, ce scénario décrit des inégalités plus fortes dans la partie supérieure de la distribution (D9/D5) et dans l'ensemble (coefficient de variation), mais pas de changement dans la partie inférieure de la

22 Cette décomposition plus fine est obtenue en autorisant à varier d'une part les inégalités entre hommes et d'autre part celles entre femmes.

distribution (D5/D2). En effet, l'association est plus forte au sommet de la distribution. Au contraire, dans sa moitié inférieure, elle est bien moins marquée ; en particulier, jusqu'au salaire médian, on n'observe guère de différence entre femmes en emploi du point de vue du salaire touché par leur conjoint.

Ce scénario prévoit un affaiblissement un peu moins marqué des inégalités d'ensemble (coefficient de variation) entre 1990 et 2011. En effet, on a déjà expliqué que l'existence d'une corrélation entre salaires des conjoints implique que l'augmentation du taux d'emploi des femmes n'aura pas un effet négatif aussi fort sur les inégalités entre couples que si ces derniers se formaient complètement au hasard. Néanmoins, cet effet d'atténuation, dont on aurait pu craindre qu'il annulerait complètement la tendance à l'égalisation due à l'augmentation de l'activité des femmes, reste d'ampleur très limitée – et même nulle en ce qui concerne les rapports interdéciles. L'effet contre-intuitif de l'augmentation de l'activité des femmes sur les inégalités s'explique donc bien par la faiblesse relative de l'association entre salaires des conjoints : l'ajout d'un second salaire à une proportion importante des couples aboutit à un arasement des écarts extrêmes, simplement parce l'association entre salaires des conjoints n'est pas assez forte pour rendre la conjonction de deux salaires très élevés ou très faibles assez probable. La mise en commun des revenus a généralement tendance à diminuer les inégalités.

La cinquième situation contrefactuelle autorise l'association entre inactivité de la femme et salaire de son conjoint à évoluer au cours du temps. Le modèle 4 qui lui correspond n'apporte aucune amélioration au modèle de stabilité (modèle 3) : l'indice de dissimilarité diminue très peu, et les critères BIC et AIC augmentent même²³. De fait, les inégalités prédites par ce modèle sont presque identiques à celles prédites par le précédent. Ainsi, comme on l'avait anticipé du fait de l'absence d'évolution de l'échelle d'activité des femmes selon le salaire de leur conjoint (Figure 3.3), ce facteur n'a joué aucun rôle dans l'évolution des inégalités entre couples. La France diffère en cela nettement des États-Unis.

La dernière situation correspond à la table réellement observée, et autorise donc finalement l'association entre salaires des conjoints à varier au cours du temps. Elle nous permet d'évaluer si ce facteur a contribué, dans un sens ou dans l'autre, à l'évolution des

23 Une spécification plus complexe visant à reproduire la forme en U inversé de l'association en introduisant une interaction avec le carré du score de salaire du conjoint n'apporte aucune amélioration.

inégalités. Par rapport à la situation précédente, les différences sont limitées. Néanmoins, on peut remarquer une diminution légèrement plus forte des inégalités d'ensemble (coefficient de variation) et surtout de celles provenant de la partie supérieure de la distribution (rapport D9/D5) entre 1990 et 2011. Ainsi, contrairement à ce qui a été relevé aux États-Unis, l'évolution de l'association entre salaires des conjoints parmi les couples dans lesquels la femme est active a plutôt joué dans le sens d'une atténuation des inégalités.

Au total, si la corrélation entre salaires des conjoints parmi l'ensemble des couples est restée très stable, voire s'est légèrement accrue (Figure 3.2, p. 198), cette tendance globale cache deux effets divergents : d'une part, une augmentation de la proportion de femmes en emploi, qui tend à renforcer l'association entre salaires des conjoints parmi l'ensemble des couples ; d'autre part, une diminution de cette association parmi les couples dans lesquels la femme est active. La progression uniforme selon les groupes sociaux du taux d'emploi des femmes a limité l'ampleur du premier mouvement, ce qui a permis au second de compenser, et même d'inverser ses effets sur l'association entre salaires des conjoints. C'est vers l'analyse de cette seconde composante que nous nous tournons maintenant.

4 L'évolution de l'association entre salaires des conjoints et de l'hypergamie au sein des couples bi-actifs

Nous venons de montrer que l'évolution de l'association entre salaires des conjoints parmi les couples dans lesquels la femme est en emploi n'avait eu qu'un effet relativement faible sur l'évolution des inégalités entre couples. Néanmoins, cet effet est légèrement négatif, ce qui implique que l'association a légèrement diminué au cours du temps. Dans cette section, nous analysons plus en détail cette association afin de déterminer non plus quelles sont les conséquences de son évolution, mais quelle est l'ampleur de cette dernière, et quelle interprétation nous pouvons en tirer concernant l'évolution de l'homogamie et de l'hypergamie.

Il s'agit d'une part d'examiner si la baisse de l'homogamie d'éducation, de classe et d'origine sociales mise en évidence au chapitre précédent trouve son parallèle dans une baisse de l'association entre salaires des conjoints parmi les couples bi-actifs. Ainsi que d'évaluer, d'autre part, si l'augmentation très nette de l'activité des femmes, et surtout de leur contribution au total des salaires des couples, s'est accompagnée d'une évolution de la

relation entre salaire de l'homme et salaire de la femme au sein de ces couples. Nous avons déjà vu plus haut un signe possible d'une telle évolution : l'augmentation de la proportion de femmes touchant un salaire supérieur à celui de leur conjoint (Figure 3.1, p. 195).

En nous concentrant ici sur les couples dans lesquels les deux conjoints touchent un salaire, nous cherchons à identifier des évolutions allant au-delà de l'effet égalisateur – entre couples autant qu'entre hommes et femmes – du développement de l'activité féminine déjà mis en évidence plus haut. Cette étude de l'hypergamie relative – à structure de la population contrôlée – doit nous permettre d'accéder à l'effet des normes sociales et des préférences individuelles sur l'asymétrie entre hommes et femmes du point de vue du salaire au sein du couple, afin d'évaluer si l'augmentation de l'activité féminine s'est accompagnée d'une véritable évolution des rôles de genre dans le couple, ou si ces derniers ont au contraire opposé une résistance aux évolutions réalisées sur le marché du travail.

4.1 La spécification de l'association entre salaires des conjoints

L'analyse de l'association entre salaires des conjoints suit nécessairement une démarche assez différente de celle retenue dans les chapitres précédents pour l'étude d'autres dimensions de l'homogamie. Du fait de sa nature quantitative, le salaire requiert de nouvelles modélisations, et autorise au moins deux lectures différentes de l'association : positionnelle ou monétaire²⁴.

Une première lecture, *positionnelle*, est possible : elle consiste à évaluer dans quelle mesure les hommes disposant des salaires les plus élevés (respectivement, les moins élevés) ont tendance à être en couple avec les femmes dans la même situation. La position de chaque conjoint est évaluée par rapport aux individus du même sexe : cette lecture met donc l'accent sur le rang social de chaque conjoint, dont le salaire peut être un indice ou du moins l'une des dimensions. On aperçoit immédiatement les limites d'une telle interprétation appliquée au salaire : comme on l'a déjà souligné, les femmes qui travaillent à temps partiel et touchent un salaire réduit peuvent le faire par contrainte (leur salaire correspond alors au revenu maximal qu'elles peuvent retirer de leur activité) ou par un choix influencé par les rôles de genre (leur salaire pourrait être supérieur, mais les revenus élevés de leur conjoint

24 C. Schwartz (2010) parle d'approches relative et absolue ; nous préférons éviter ces termes, qui prêtent ici à confusion avec l'opposition entre homogamie absolue et homogamie relative.

leur permettent de limiter leur investissement sur le marché du travail). En outre, les salaires féminins sont aussi le résultat de choix passés concernant notamment les interruptions de carrière. Ainsi, le salaire réellement touché reflète certes un niveau de qualification et plus largement le statut social, mais celui-ci entre en interaction avec les rôles de genre, et peut s'en trouver passablement brouillé.

Une seconde lecture, *monétaire*, revient à l'inverse à retenir la valeur concrète des salaires : elle correspond à l'idée qu'une somme présente la même valeur pour le couple quel que soit le conjoint qui la touche, ou que les conjoints peuvent être amenés à comparer leurs salaires. Cette lecture se concentre ainsi sur la valeur économique du salaire. Néanmoins, elle est aussi d'un grand intérêt pour l'étude des rôles de genre : en effet, c'est au niveau du couple que la comparaison entre contributions des conjoints fait sens pour les individus, et non pas au niveau plus abstrait de l'ensemble de la population (qui est le niveau retenu par l'approche positionnelle). Dans cette logique, l'augmentation de la participation des femmes au marché du travail pourrait avoir transformé la signification sociale de leur salaire, qui ne serait plus considéré comme un revenu d'appoint (Schwartz, 2010, p. 1532). En effet, en amont de la mise en couple, une vision plus fréquemment égalitaire du couple pourrait inciter les individus à rechercher des conjoints de niveau social proche en termes de revenus ; en outre, on peut supposer qu'un pouvoir d'achat similaire est associé à une similarité de styles de vie qui peut favoriser la rencontre et la mise en couple. En aval de la mise en couple, les conjoints pourront avoir tendance à adapter leurs attentes et leurs comportements sur le marché du travail de manière à limiter leur écart de salaire, qui est en effet une source de pouvoir et un élément de négociation dans la définition des rôles et la répartition du travail domestique, mais aussi une cause potentielle de tensions dans un contexte de remise en cause des rôles de genre (de Singly et Glaude, 1986 ; Testenoire, 2008). Cette hypothèse nous semble en tout cas devoir être confrontée aux données empiriques.

Ces deux approches trouvent plusieurs traductions possibles dans des spécifications de la structure de l'association dont il est possible de tester la pertinence en comparant les statistiques d'ajustement des trois modèles. Nous présentons tout d'abord trois spécifications de l'association symétrique (dans laquelle les positions des conjoints sont

interchangeables), puis leur équivalent concernant la modélisation de l'hypergamie (association asymétrique)²⁵.

4.1.1 L'association symétrique

La première spécification symétrique de l'association, notée (a), est l'équivalent direct du coefficient de corrélation de Pearson utilisé plus haut. Elle est obtenue à l'aide d'une interaction linéaire-par-linéaire entre les salaires des conjoints, après calcul de scores standardisés (cf. Encadré 3.2). Cette spécification correspond à l'approche positionnelle, puisque le logarithme de l'*odds ratio* généralisé est ici inversement proportionnel au produit des écarts de salaire à la moyenne de chaque sexe. Néanmoins, la valeur monétaire du salaire compte : c'est bien l'ampleur de l'écart à la moyenne qui est déterminant.

Toujours suivant l'approche positionnelle, une deuxième spécification, notée (b), consiste à considérer que seul le rang dans la hiérarchie des salaires de son sexe compte. Dans ce modèle (dit de distance, cf. Encadré 3.2), le logarithme de l'*odds ratio* généralisé est alors inversement proportionnel à la valeur absolue (au sens ici de symétrique entre les deux conjoints) de l'écart entre conjoints en nombre de quantiles de salaire. La lecture en termes de rang social est ici encore plus nette : les individus choisiraient leur conjoint en respectant le classement suivant leur salaire des individus du sexe opposé. L'écart monétaire ne joue ici plus aucun rôle.

Finalement, l'approche monétaire peut être mise en œuvre par un autre modèle de distance, noté (c), dans lequel le logarithme de l'*odds ratio* généralisé est inversement proportionnel à la valeur absolue de l'écart entre salaires des conjoints exprimés en euros constants 2011.

4.1.2 L'hypergamie

Les trois spécifications que nous venons de décrire sont totalement aveugles à l'hypergamie féminine, dans la mesure où les rôles des deux conjoints y sont parfaitement interchangeables. Mais les deux dernières, fondées sur des modèles de distance, peuvent être étendues de manière à tenir compte de l'asymétrie hommes-femmes. Il suffit pour cela d'introduire en plus de la valeur absolue de l'écart, un paramètre indiquant si l'homme

25 Nous avons testé d'autres modèles, comme le modèle UNIDIFF ou le modèle à forme de régression, utilisés au chapitre précédent. Aucun ne s'ajuste mieux aux données que ceux retenus ici.

occupe une position supérieure, en termes positionnels ou monétaires, à celle de sa conjointe (Mare, 1991, p. 22). Ce paramètre prendra une valeur positive en cas d'existence d'une hypergamie féminine au-delà de la contrainte des marges de la table. Ce résultat signifierait alors qu'étant donné la structure des salaires chez les deux sexes, les couples se forment et ajustent leur participation sur le marché du travail de manière à éviter que la femme gagne plus que son conjoint.

Le type d'approche – positionnelle ou monétaire – retenu influe de manière fondamentale sur le sens de l'hypergamie mesurée. Dans l'approche positionnelle, l'hypergamie féminine impliquerait que les femmes prennent généralement pour conjoints des hommes mieux placés qu'elles en termes de rang des salaires de chaque sexe considéré séparément. Le paramètre introduit dans le modèle identifie donc les configurations dans lesquelles l'homme appartient à un quantile supérieur à celui de sa conjointe. Dans l'approche monétaire, l'hypergamie féminine suppose qu'au sein du couple, l'homme gagne généralement plus que sa conjointe, et ce sont ces configurations qui sont identifiées par le paramètre.

Encadré 3.2 : Équations des modèles log-linéaires utilisés pour l'analyse de l'association

Les modèles utilisés dans cette section se fondent tous sur le modèle d'indépendance conditionnelle (modèle 1) défini à la section précédente (*cf.* Encadré 3.1, p. 207). Chaque spécification introduite consomme un degré de liberté supplémentaire, que ce soit pour la description de l'association, de l'hypergamie, ou de leur évolution temporelle.

Trois spécifications décrivent chacune une manière différente de tenir compte de l'association symétrique entre salaires des conjoints. La spécification (a) la décrit sous la forme d'une association linéaire-par-linéaire (Haberman, 1974 ; Agresti, [1990] 2002, sect. 9.4.1), dite aussi association uniforme généralisée (Goodman, 1985, 1986, 1991). Ce modèle, qui est similaire au modèle RC utilisé au chapitre précédent mais avec des scores fixés *a priori*, présente l'intérêt de fournir un coefficient d'association ϕ équivalent (à une transformation près) au coefficient de corrélation de Pearson lorsque les scores sont standardisés et que les données suivent une distribution normale bivariée. Ces scores de salaire standardisés correspondent simplement aux salaires médians des quantiles

d'appartenance des individus (une catégorie séparée correspond aux inactifs et aux chômeurs), auxquels on impose une contrainte de centrage et de réduction. Le modèle s'écrit donc dans sa version stable dans le temps :

$$M5a : MI + \phi s_h^H s_f^F$$

et sa version avec évolution linéaire au cours du temps :

$$M6a : M5a + t\phi^T s_h^H s_f^F$$

avec s_h^H et s_f^F respectivement les scores de salaire de l'homme et de la femme, et t l'année d'enquête (variable numérique continue). L'*odds ratio* conditionnel pour l'année t est donc égal à :

$$\theta_{hf, h'f'}(t) = \exp \left[(\phi + t\phi^T)(s_h^H - s_{h'}^H)(s_f^F - s_{f'}^F) \right]$$

Les deux autres spécifications correspondent à des modèles de distance (Haberman, 1974 ; Goodman, 1979). La version (b), positionnelle, fait dépendre la sur- ou sous-représentation de la configuration de couple de l'écart absolu entre numéros des quantiles des conjoints. Le modèle s'écrit dans sa version stable :

$$M5b : MI + \lambda^P |h - f|$$

et sa version avec évolution temporelle linéaire :

$$M6b : M5b + t\lambda^{PT} |h - f|$$

L'*odds ratio* conditionnel dépend de l'effet général de la distance une année donnée (dit paramètre de pas), que multiplie la différence entre la somme des écarts absolus entre conjoints dans les deux cellules comparées et celle des écarts absolus entre conjoints dans les deux autres cellules entrant dans le calcul de l'*odds ratio* :

$$\theta_{hf, h'f'}(t) = \exp \left[(\lambda^P + t\lambda^{PT}) (|h - f| + |h' - f'| - |h - f'| - |h' - f|) \right]$$

En particulier, on peut remarquer que les *odds ratios* contrastant deux cellules de la diagonale sont inversement proportionnels (sur l'échelle logarithmique) à l'écart absolu entre ces deux cellules.

Dans la version (c), monétaire, la sur- ou sous-représentation dépend de l'écart absolu de salaire entre conjoints en euros. Le modèle s'écrit dans sa version stable :

$$M5c : MI + \lambda^M |m_h - m_f|$$

et sa version avec évolution temporelle linéaire :

$$M6c : M5c + t\lambda^{MT} |m_h - m_f|$$

avec m_h et m_f respectivement les salaires en euros constants 2011 de l'homme et de la femme. L'*odds ratio* conditionnel est similaire à celui de la spécification précédente :

$$\theta_{hf, h'f'(t)} = \exp \left[(\lambda^M + t\lambda^{MT}) (|m_h - m_f| + |m_{h'} - m_{f'}| - |m_h - m_{f'}| - |m_{h'} - m_f|) \right]$$

Enfin, la spécification (bc) combine les deux dernières versions. Soit dans sa version stable :

$$M5bc : MI + \lambda^P |h - f| + \lambda^M |m_h - m_f|$$

et sa version avec évolution temporelle linéaire :

$$M6bc : M5bc + t\lambda^{PT} |h - f| + t\lambda^{MT} |m_h - m_f|$$

Les spécifications de l'hypergamie se fondent sur le signe des distances que nous venons de présenter. Les paramètres s'appliquent uniquement aux couples dans lesquels l'homme occupe une position supérieure à sa conjointe, selon les deux définitions retenues ici. L'effet est donc exprimé en comparaison avec la fréquence des couples dans lesquels la femme a une position supérieure ou égale à celle de son conjoint.

Les *odds ratios* construits uniquement à partir de configurations dans lesquelles l'homme (respectivement, la femme) touche le salaire le plus élevé ne sont donc pas affectés. Seuls le sont ceux contrastant des configurations entre lesquelles la relation entre les salaires des conjoints s'inverse : les *odds ratios* contrastant une configuration dans laquelle l'homme touche le salaire le plus élevé avec une configuration dans laquelle ce n'est pas le cas sont tous augmentés de manière uniforme par un coefficient multiplicatif ; ceux construits de manière inverse sont au contraire diminués dans la même proportion.

Le modèle incluant une spécification positionnelle pour les associations symétrique et asymétrique s'écrit :

$$M7b/b : M6b + \lambda^{P'} \mathbb{1}_{h>f}$$

et sa version avec évolution temporelle linéaire :

$$M8b/b : M7b/b + t\lambda^{P'T} \mathbb{1}_{h>f}$$

Le modèle incluant une spécification monétaire s'écrit :

$$M7c/c : M6c + \lambda^{M'} \mathbb{1}_{m_h>m_f}$$

et sa version avec évolution temporelle linéaire :

$$M8c/c : M7c/c + t\lambda^{M'T} \mathbb{1}_{m_h>m_f}$$

On laissera au lecteur le soin de combiner le modèle 6bc avec les spécifications introduites dans les modèles 7b/b et 7c/c afin d'obtenir les modèles 7bc/b, 7bc/c et 7bc/bc, ainsi qu'avec celles introduites dans ces deux modèles et les modèles 8b/b et 8c/c afin d'obtenir les modèles 8bc/b, 8bc/c, 8bc/bc et 9bc/bc.

4.2 Résultats : affaiblissement de l'association, résistance de l'hypergamie

4.2.1 Ajustement des modèles

Nous commençons par tester les différentes spécifications de l'association symétrique, d'abord de manière stable, puis en autorisant des évolutions linéaires au cours du temps. Puis nous introduisons différentes spécifications pour l'hypergamie, d'abord stables, puis variables.

Nous repartons du modèle d'indépendance (modèle 1 dans la section précédente), à partir d'une table d'homogamie différente puisque les couples dans lesquels l'un des deux conjoints ne touche aucun salaire ont été exclus²⁶. Ce modèle de base classe de manière incorrecte 20 % des couples (Tableau 3.4). Les trois spécifications concurrentes de l'association entre salaires des conjoints (modèles 5a, 5b et 5c) s'ajustent de manière assez similaire aux données, en faisant tomber cette proportion entre 13 et 15 %. Néanmoins, on peut remarquer que la troisième spécification décrit nettement mieux les données que les

26 Ceci revient à retirer la première ligne et la première colonne de la table précédente.

deux autres, pour un même nombre de degrés de liberté. En complexifiant les modèles, on observe que la combinaison des deux dernières spécifications offre la meilleure description d'après la déviance, et donc selon les indicateurs BIC et AIC (modèle 5bc)²⁷. Ce résultat rejoint une observation de C. Schwartz (2010) sur des données américaines.

Les deux logiques, à la fois positionnelle et monétaire, semblent donc expliquer l'association observée entre salaires des conjoints. Ainsi, le coefficient de corrélation, auquel nous avons eu recours plus haut, n'apparaît pas comme le meilleur indicateur de l'association entre salaires des conjoints : l'analyse détaillée à l'aide de modèles de distance est en mesure d'améliorer la compréhension du phénomène.

L'évolution temporelle de l'association ne fait aucun doute : le modèle autorisant les deux spécifications à varier séparément au cours du temps (modèle 6bc) est nettement préféré par le BIC et l'AIC. Le gain en termes de description des données est cependant relativement faible, ce qui n'étonne guère au vu de la faible influence de l'évolution de l'association sur les inégalités déjà mise en évidence plus haut : 0,07 points de pourcentage seulement²⁸.

À partir de ce modèle, l'existence d'une asymétrie entre hommes et femmes est elle aussi particulièrement claire : quelle que soit la spécification de l'hypergamie retenue, le BIC et l'AIC diminuent très fortement, même si l'indice de dissimilarité ne baisse que de 0,1 à 0,2 points. Les spécifications positionnelle (modèle 7bc/b) et monétaire (7bc/c) sont de qualités très proches, quoique la seconde soit très légèrement meilleure. Le meilleur modèle (7bc/bc) inclut les deux spécifications : ceci semble indiquer que les deux logiques sont valables.

27 L'indice de dissimilarité est plus ambigu. Nous préférons suivre la déviance, qui constitue à notre sens un indicateur plus fiable dans le cadre de tables aussi détaillées que celle utilisée ici.

28 Soulignons encore une fois que la table utilisée ici est très détaillée, ce qui tend à rendre difficile une diminution de l'indice de dissimilarité.

Tableau 3.4 : Indicateurs d'ajustement des modèles d'analyse de l'association

	D. L.	Déviante	Δ (%)	BIC	AIC
M1 : Indépendance cond.	7 942	29 242	19,74	-62 804	13 358
M5a	7 941	15 365	14,37	-76 669	-517
M5b	7 941	15 744	14,72	-76 290	-138
M5c	7 941	11 743	12,67	-80 291	-4 139
M5ab	7 940	13 502	13,51	-78 520	-2 378
M5ac	7 940	11 299	12,30	-80 723	-4 581
M5bc	7 940	11 156	12,31	-80 866	-4 724
M6a	7 940	15 335	14,36	-76 687	-545
M6b	7 940	15 715	14,70	-76 308	-165
M6c	7 940	11 622	12,63	-80 401	-4 258
M6bc	7 938	11 038	12,24	-80 961	-4 838
M7b/b	7 939	15 107	14,47	-76 904	-771
M7c/c	7 939	11 273	12,42	-80 738	-4 605
M7bc/b	7 937	11 014	12,23	-80 974	-4 860
M7bc/c	7 937	10 994	12,22	-80 994	-4 880
M7bc/bc	7 936	10 957	12,21	-81 019	-4 915
M8b/b	7 939	15 136	14,49	-76 874	-742
M8c/c	7 939	11 342	12,44	-80 668	-4 536
M8bc/b	7 936	11 014	12,23	-80 962	-4 858
M8bc/c	7 936	10 986	12,21	-80 990	-4 886
M8bc/bc	7 934	10 950	12,20	-81 003	-4 918
M9bc/bc	7 935	10 950	12,20	-81 015	-4 920

D. L. : degrés de liberté. Δ : indice de dissimilarité.

Finalement, nous observons des signes contradictoires concernant l'existence d'une évolution temporelle de l'hypergamie. L'introduction d'une tendance temporelle linéaire n'améliore absolument pas la description des données concernant la spécification positionnelle (modèle 8bc/b), mais elle l'améliore légèrement concernant la spécification monétaire (modèle 8bc/c). La prise en compte des deux spécifications avec des tendances temporelles linéaires fournit un modèle (modèle 8bc/bc) encore meilleur d'après l'AIC, mais le BIC préfère le modèle de stabilité (modèle 7bc/bc). En fait, un modèle n'autorisant que la

spécification monétaire de l'hypergamie à varier dans le temps (modèle 9bc/bc) est encore meilleur. Il semble donc que seule l'hypergamie en termes monétaires ait varié – mais pas de manière radicale.

Afin de mieux comprendre l'interaction entre les deux spécifications différentes de la distance et de l'hypergamie qui sont combinées dans la spécification (bc/bc), nous présentons aussi ci-dessous les résultats obtenus séparément pour chaque spécification. Ici encore, l'évolution temporelle est significative (les modèles 6b et 6c constituent une amélioration par rapport aux modèles 5b et 5c), de même que l'hypergamie (modèles 7b/b et 7c/c). En revanche, la prise en compte de l'évolution temporelle de l'hypergamie (modèles 8b/b et 8c/c) n'améliore pas ces modèles avec spécifications séparées, contrairement à ce que l'on observait lorsque les deux spécifications étaient prises en compte simultanément (modèle 8bc/bc). Comme nous allons le voir, ce résultat étonnant tient au fait que l'évolution de l'hypergamie selon l'approche monétaire s'explique par l'évolution du lien entre les deux spécifications de l'association symétrique.

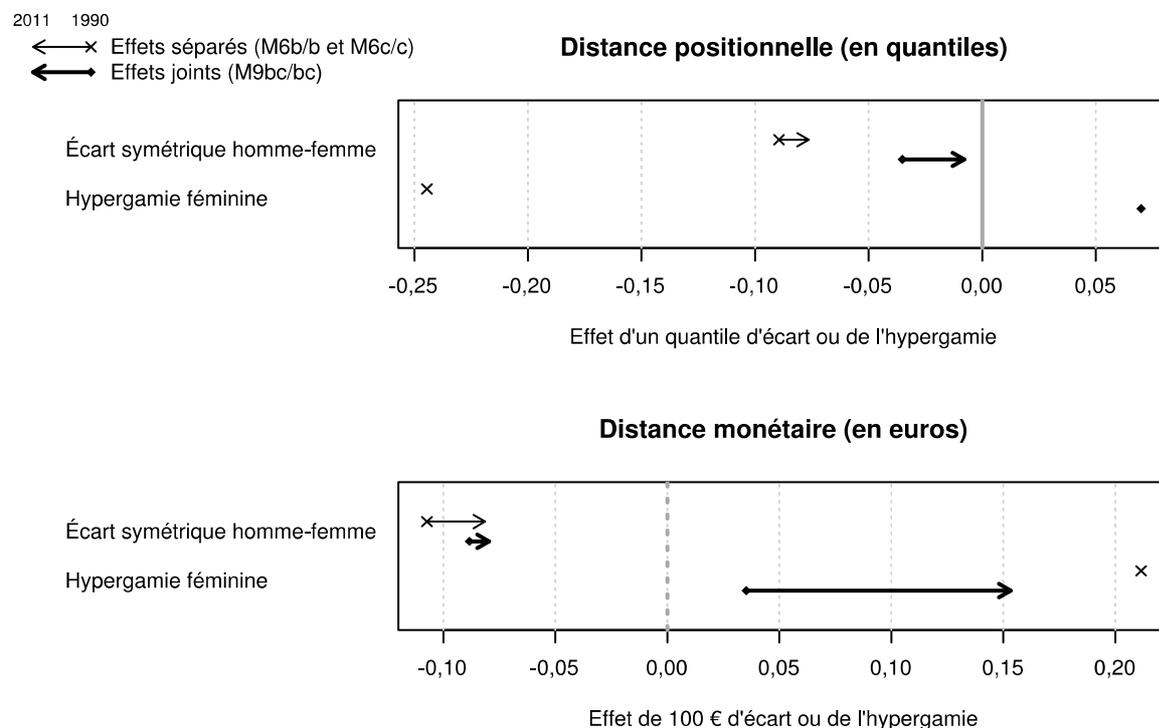
4.2.2 Un affaiblissement de l'association en termes positionnels et monétaires

La figure 3.6 présente les coefficients de distance et d'hypergamie obtenus à partir des approches positionnelle (graphique du haut) et monétaire (graphique du bas). La première série de coefficients correspond aux modèles intégrant une spécification à la fois (modèles 6b/b et 6c/c), alors que la seconde série présente ceux du modèle à double spécification (modèle 9bc/bc)²⁹. Rappelons qu'au vu de l'ajustement des modèles, l'évolution temporelle de l'hypergamie n'a été retenue que pour le modèle complet, et dans sa spécification monétaire uniquement.

Les coefficients issus des modèles à effets séparés mettent tous les deux en évidence, comme attendu, un effet négatif de l'écart entre salaires des conjoints sur les chances de rencontrer une configuration de couple donnée, au sein des couples bi-actifs. Néanmoins, on note un affaiblissement assez important de cette association symétrique entre 1990 et 2011, de l'ordre de 15 % pour la spécification positionnelle et de 25 % pour la monétaire.

29 Sauf mention contraire, tous les effets commentés sont très fortement significatifs.

Figure 3.6 : Évolution de l'association symétrique et de l'hypergamie parmi les couples bi-actifs entre 1990 et 2011



Source : enquêtes Emploi (Insee)

Champ : hommes et femmes vivant en couple cohabitant dans lequel l'un des conjoints a entre 30 et 59 ans, les deux conjoints sont en emploi et aucun n'est indépendant

Le modèle avec effets joints (M9bc/bc) n'autorise pas l'effet l'hypergamie en termes positionnels à varier au cours du temps. Les modèles avec effets séparés (M6b/b et M6c/c) n'autorisent aucun effet de l'hypergamie à varier. Dans ces trois cas, le coefficient correspondant apparaît comme un point unique.

L'introduction simultanée des deux spécifications (modèle 9bc/bc) réduit l'ampleur de chacun des effets sans modifier les conclusions : une partie de l'association mesurée précédemment par chaque spécification est capturée par l'autre, la hiérarchie des quantiles étant naturellement assez fortement corrélée aux niveaux de salaire. C'est l'association positionnelle qui diminue le plus, ce qui, avec le fait que la spécification monétaire décrit mieux les données, confirme que c'est l'écart de salaire en termes monétaires qui joue le plus fort rôle dans l'appariement des conjoints, plus que le rang social en référence à chaque sexe – qui est une notion plus abstraite. Concernant l'évolution temporelle on peut noter que la baisse de l'association positionnelle triple (à 78 %), alors que celle de l'association

monétaire est au contraire atténuée (à 10 %). Cette quasi-disparition de l'association positionnelle semble signaler une plus grande commensurabilité des salaires des hommes et des femmes : la comparaison entre conjoints nécessite de moins en moins de prendre pour référence la moyenne de chaque sexe, mais peut se faire directement en termes monétaires. On peut tenter d'expliquer ce mouvement par l'augmentation – modeste – des salaires relatifs des femmes en emploi par rapport à ceux des hommes, mise en évidence dans la première section, qui, associée à des carrières féminines moins souvent interrompues, a pu entraîner un changement de comportement des couples (rappelons que l'évolution de la distribution marginale en termes de salaires est ici contrôlée).

Les coefficients mesurant l'association asymétrique offrent un résultat plus étonnant. On observe bien que les chances de rencontrer une configuration donnée sont nettement plus fortes lorsque l'homme touche le salaire le plus élevé du couple, en termes monétaires (modèle 6c/c). En revanche, c'est l'inverse qui est vrai lorsque seule l'approche positionnelle est retenue (modèle 6b/b) : les cas dans lesquels la femme appartient à un quantile de salaire supérieur à celui de l'homme sont sur-représentés. Les deux spécifications mesurent des effets d'ampleurs comparables, mais de sens opposés³⁰. On voit donc que le décalage entre le salaire médian des hommes et celui des femmes produit un chiasme concernant la relation entre salaires des conjoints : si les hommes gagnent généralement plus que leur conjointe, ils ont en revanche tendance à être en couple avec des femmes placées plus haut qu'eux dans la hiérarchie des salaires propre à chaque sexe.

Comme pour l'association symétrique, la prise en compte simultanée des deux spécifications (modèle 9bc/bc) réduit très nettement l'hypergamie selon les deux définitions³¹. Mais surtout, elle inverse ici le sens de l'hypergamie dans la spécification positionnelle : dans le modèle complet, l'hypergamie féminine prédomine suivant les deux spécifications. L'interprétation avancée au paragraphe précédent est donc confirmée : l'hypogamie féminine en termes de position dans l'échelle des salaires propre à chaque sexe disparaît dès que l'on tient compte de l'association en termes monétaires.

30 Ces variables étant qualitatives, il est possible de comparer directement la valeur des coefficients, contrairement à l'effet d'un quantile ou d'un euro d'écart, qui doivent par exemple être rapportés aux écarts-types.

31 L'effet de l'hypergamie en termes monétaires est même non significatif en 1990.

Néanmoins, dans le modèle complet, alors que la spécification positionnelle identifie une stabilité de l'hypergamie, la spécification monétaire mesure, elle, un triplement de l'effet entre 1990 et 2011. Il est difficile de ne pas rapprocher ce mouvement de celui, d'ampleur comparable mais de sens opposé, suivi par l'association symétrique en termes positionnels. De fait, on a vu plus haut que l'évolution temporelle de l'hypergamie n'était significative que dans un modèle tenant compte des deux manières de spécifier l'association symétrique. Ce phénomène s'explique par le fait que l'hypergamie féminine en termes monétaires était auparavant en partie masquée par une association symétrique en termes positionnels : les salaires des femmes étant inférieurs à ceux des hommes, une proximité en termes positionnels (appartenance à des quantiles proches) revient à une hypergamie féminine en termes monétaires (salaire de la femme inférieur à celui de son conjoint), et vice-versa. La réduction de l'écart entre salaires masculins et féminins révèle donc progressivement, dans ce modèle, l'existence – et la résistance – d'une hypergamie féminine en termes monétaires dont on voyait avec le modèle à effets séparés (modèle 6c/c) qu'elle n'a pas diminué. Si les salaires féminins devaient finir par rattraper les salaires masculins – ce qui, au vu des tendances déjà évoquées, ne paraît pas assuré –, on peut penser que l'effet convergerait vers celui mesuré dans ce dernier modèle.

Au total, les analyses réalisées dans cette section confirment que l'association entre salaires des conjoints a diminué parmi les couples bi-actifs. Cet affaiblissement, de l'ordre de 15 à 25 %, s'observe aussi bien en retenant une approche positionnelle comparant les positions de chacun des conjoints dans la hiérarchie des salaires de leur sexe, qu'en adoptant une lecture en termes d'écart monétaires. Cependant, l'évolution a été plus marquée en termes positionnels, si bien que cette lecture a presque complètement perdu de sa pertinence. Ce mouvement semble bien plus net que la modeste augmentation des salaires des femmes par rapport à ceux des hommes. Il semble donc refléter aussi une modification des comportements, autant du point de vue du choix du conjoint que des décisions d'activité prises après la mise en couple. De ce point de vue, l'hypothèse proposée par C. Schwartz paraît validée.

En revanche, ce mouvement qui aurait pu traduire une plus grande égalité des sexes ne paraît pas avoir entamé la tendance à l'hypergamie féminine par le salaire : celle-ci apparaît

très stable selon les deux lectures – positionnelle et monétaire – lorsqu’elles sont adoptées séparément. Soulignons cependant que si l’on observe bien une hypergamie féminine en termes monétaires, on note une hypogamie féminine, moins attendue, en termes de quantile de salaire des conjoints, qui est liée à l’infériorité des salaires des femmes par rapport à ceux des hommes. Mais ce phénomène peut n’être considéré que comme un artefact, dans la mesure où il disparaît lorsque l’on tient compte des salaires en termes monétaires. Si l’on considère que l’hypergamie en termes monétaires est la plus significative, du fait de ses conséquences financières, et de son influence sur les rapports de pouvoir au sein du couple, alors force est de constater que ce phénomène est loin de disparaître. L’augmentation des salaires relatifs féminins n’a pour l’instant pas suffi à aller à l’encontre de la tendance à l’hypergamie : l’affaiblissement de l’association entre quantiles de salaire des conjoints a révélé de manière encore plus nette la préférence pour l’hypergamie féminine en termes d’écarts monétaires.

Rappelons cependant, en conclusion de cette section, que ces résultats ne portent que sur les couples bi-actifs, et que nous mesurons ici la tendance à l’hypergamie au-delà de l’influence de la structure par salaire de la population. Le maintien de l’hypergamie relative revient à dire que les couples bi-actifs continuent à se former et à ajuster leur activité de manière à ce que l’homme gagne généralement plus que sa conjointe. Cependant, la diminution de la proportion des couples dans lesquels la femme ne touche aucun salaire a en elle-même réduit l’écart de salaires entre conjoints, et nous avons vu plus haut que la proportion de femmes gagnant plus que leur conjoint a nettement augmenté au cours du temps³². Les résultats que nous venons de commenter, plus exigeants, ne font que nuancer ce constat plus global, en soulignant que ces évolutions ne semblent pas avoir été suivies d’un affaiblissement parallèle de la préférence pour l’hypergamie féminine.

On ne peut cependant exclure que la stabilité observée ici ne soit en fait le produit d’une plus forte tendance à l’hypergamie féminine des couples dans lesquels la femme était précédemment inactive : dans ce cas, on ne pourrait exclure qu’un affaiblissement de l’hypergamie soit cependant à l’œuvre parmi les couples bi-actifs de plus longue date. Cet affaiblissement deviendrait plus visible à mesure que la progression de la proportion de

32 Cette évolution se retrouve parmi les couples bi-actifs, avec une proportion de femmes gagnant plus que leur conjoint supérieure d’environ 10 points de pourcentage à celle observée parmi l’ensemble des couples.

couples bi-actifs, ayant atteint un plafond, ralentit, et que les couples adoptent progressivement un comportement plus égalitaire. Mais rien ne nous permet à ce stade d'affirmer qu'un tel scénario est à l'œuvre.

5 Une stabilité de l'association entre statuts d'emploi des conjoints

L'analyse de l'évolution de l'association entre salaires des conjoints offre un complément important à celle de l'homogamie d'éducation, de classe ou d'origine sociales. Néanmoins, le salaire – ou même plus largement le revenu – ne représente qu'une des dimensions de la position sociale des conjoints en termes économiques. Tout aussi importante apparaît la question de la précarité de l'emploi, qui recouvre, en plus de la question de l'intégration sociale par le travail, celle de la stabilité des revenus. Concernant ce second aspect, qui est celui sur lequel nous avons mis l'accent ici, certains travaux (Hyslop, 2001) ont pu distinguer, dans une approche longitudinale, entre composantes instantanée et permanente du revenu, de manière à prendre en compte les fluctuations qui peuvent affecter la mesure du lien entre salaires des conjoints. Il apparaît que la corrélation entre salaires des conjoints tient presque exclusivement à la composante stable (« permanente ») du revenu : de ce fait, l'activité des deux conjoints offre une certaine stabilisation des revenus du couple par rapport à ce que toucheraient les conjoints séparément. Ici encore, le développement de l'emploi féminin a pu constituer un facteur d'égalisation entre couples.

Néanmoins, l'accroissement du chômage et de la précarité de l'emploi qui ont eu lieu dans le même temps ont pu contrarier cette tendance. En effet, la précarité n'affecte pas seulement, comme à l'évidence, la stabilité des revenus individuels (Breuil-Genier, 2001) : puisque celle-ci touche plus fortement certaines catégories sociales, la tendance à l'homogamie sociale implique que les couples associeront plus fréquemment que le hasard ne le voudrait des conjoints souffrant de la précarité de l'emploi (Ultee, Dessens et Jansen, 1988)³³. C'est pourquoi nous concluons ce chapitre par une étude de l'association entre les

33 La précarité peut aussi avoir des effets sur la vie familiale, et notamment entraîner des difficultés à se mettre en couple ou des séparations (ce dernier point étant discuté, voir Paugam, 2000, chap. 8 ; Solaz, 2013), ce qui pourrait avoir pour conséquence de diminuer la prévalence des couples comprenant au moins un individu en situation de précarité.

statuts d'emploi des conjoints, et son évolution depuis 1975 et 1982. Nous nous demanderons si le développement d'un chômage de masse, et le développement des contrats dits atypiques a entraîné la constitution d'un groupe social doté d'une certaine cohérence du point de vue des alliances conjugales, et sur lequel se concentreraient durablement les difficultés liées à la précarité de l'emploi.

Nous nous concentrons ici encore sur les couples bi-actifs. Bien que ce choix implique d'exclure du champ les individus qui, éprouvant le plus de difficultés à retrouver un emploi, renoncent même à se déclarer comme chômeurs, il n'est pas possible de toujours considérer l'inactivité comme indicateur d'une situation précaire. Comme on l'a vu plus haut, l'inactivité féminine peut au contraire parfois être associée à une certaine aisance financière du couple. En outre, la nette augmentation du taux d'activité féminine au cours du temps introduirait un biais très important, qui rendrait impossible toute étude des effets du développement des statuts précaires sur la période.

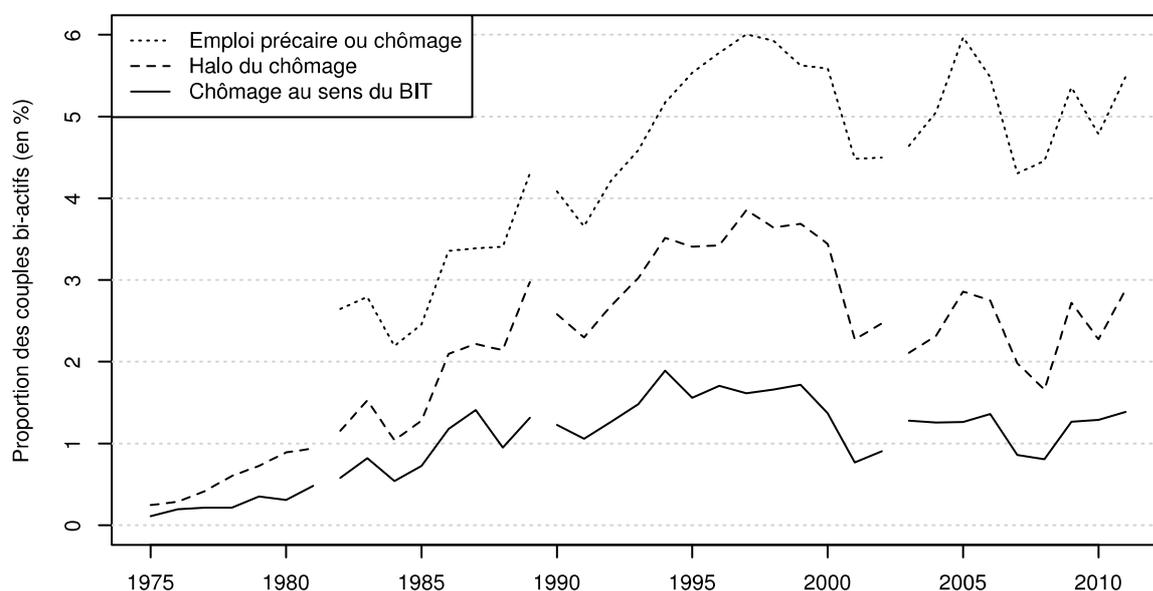
5.1 Une explosion de la proportion de couples formés de deux précaires

En guise de première approche de l'évolution de la précarité des couples, la figure 3.7 représente la proportion des couples bi-actifs dans lesquels les deux conjoints peuvent être considérés comme précaires selon les trois définitions établies plus haut : chômage au sens strict selon la définition du Bureau international du travail (BIT) ; halo du chômage, suivant la déclaration spontanée de leur statut par les interrogés, et en tenant compte de la situation un an avant l'enquête ; précarité de l'emploi au sens large. Quel que soit l'indicateur retenu, on observe une forte croissance de la proportion de couples formés de deux précaires : quasi-nulle en 1975 pour les deux définitions du chômage, pour atteindre respectivement 1,4 et 2,8 % en 2011 ; inférieure à 3 % avant 1985 pour la définition large de la précarité, pour atteindre 5,5 % en 2011.

Sans surprise, ces proportions suivent d'assez près l'évolution du chômage et de la précarité des emplois (Paugam, 2000, chap. 2 ; Insee, 2013). Les évolutions conjoncturelles sont bien marquées, et parallèles pour les trois courbes, quoique les pics du halo du chômage et de la précarité de l'emploi présentent souvent un léger retard par rapport à celles du chômage au sens du BIT. On remarque en particulier une phase d'augmentation

entre 1975 et 1994 ou 1997 (avec un pic local en 1987 ou 1989), puis une décade jusqu'en 2001 suivie d'un brusque rebond et d'une rechute avant la dernière phase de remontée à partir de 2008. Dans l'ensemble, la cohérence des séries d'enquête apparaît relativement bonne.

Figure 3.7 : Évolution de la proportion de couples formés de deux chômeurs ou précaires entre 1975 et 2011



Source : enquêtes Emploi (Insee)

Champ : hommes et femmes vivant en couple cohabitant dans lequel l'un des conjoints a entre 30 et 59 ans et les deux conjoints sont actifs.

Le chômage au sens du BIT est évalué à la date de l'enquête. Le halo du chômage désigne les individus se déclarant spontanément chômeurs à la date de l'enquête ou un an avant. La précarité comprend cette dernière catégorie, ainsi que l'ensemble des contrats dits atypiques (ni contrat à durée indéterminée, ni titulaire de la fonction publique). Cf. la section « Données » (p. 189) pour plus de précisions.

On voit donc qu'une proportion non négligeable des couples se trouvent, à un moment donné sinon de manière plus durable – ce que nous ne pouvons véritablement appréhender ici – dans une situation de précarité simultanée des deux conjoints. Cette proportion fluctue,

selon la conjoncture, entre 4,5 et 6 % dans les années récentes, selon la définition la plus extensive. Si ce constat n'est guère étonnant étant donné la croissance globale des statuts précaires, une tendance à la constitution d'un groupe social des précaires est-elle décelable ?

5.2 Une stabilité de l'endogamie relative du groupe des précaires

Pour avancer des éléments de réponse à cette question, nous reprenons ici la démarche mobilisée au chapitre précédent pour l'étude de groupes sociaux dotés d'une certaine cohérence : les classes sociales et les groupes de diplôme. En effet, on se rappelle que, dans une lecture wébérienne, l'intensité de la mise en couple à l'intérieur d'un groupe peut être considérée comme l'un des facteurs d'apparition ou de maintien d'une conscience de soi – autant qu'un symptôme d'une similarité de conditions de vie, d'une ségrégation spatiale ou de contacts fréquents. C'est l'homogamie relative qui constitue l'indice de cette tendance à l'entre-soi.

Les couples de deux précaires se forment-ils plus souvent que le hasard ne le supposerait ? Des modèles log-linéaires décrivant l'unique *odds ratio* d'une table à quatre cellules croisant chaque année le statut de chacun des conjoints suivant les trois indices de précarité répondent positivement à cette question (Tableau 3.5). Si le modèle d'indépendance décrit relativement bien les données, cela ne tient en effet qu'à la faible proportion de couples constitués de deux précaires. L'introduction d'une association stable au cours du temps améliore très nettement les trois modèles, et divise l'indice de dissimilarité par un facteur de 8 à 10.

Les *odds ratios* mesurant les chances d'endogamie augmentent à mesure que la définition de la précarité retenue est plus restrictive, passant de 2,5 pour la définition large à 3,1 pour le halo du chômage, et 3,2 pour le chômage au sens strict. Ces valeurs sont très similaires à celles obtenues pour plusieurs pays développés dans les années 1980 par une étude précédente (Ultee, Dessens et Jansen, 1988, p. 113)³⁴. Loin d'être négligeables, ces niveaux d'association restent cependant relativement faibles en comparaison de ceux mesurés au chapitre précédent (Figure 2.2, p. 151) : ils sont à peine supérieurs aux chances

34 Cette étude avait par ailleurs montré que l'association entre statuts d'emploi des conjoints ne pouvait s'expliquer simplement par l'homogamie d'éducation, la proximité des âges des conjoints, et leur appartenance au même marché du travail local. Ce résultat pourrait indiquer que les chômeurs présentent une certaine unité par-delà leurs diverses appartenances ; néanmoins, l'association reste d'ampleur limitée.

d'endogamie des classes sociales et des groupes de diplôme les plus ouverts, et sont inférieurs à ceux observés pour les ouvriers non qualifiés ou les non diplômés, les plus frappés par la précarité. Ces chiffres confirment ainsi les travaux selon lesquels les précaires ne forment pas véritablement un groupe social homogène, encore moins une classe sociale dotée d'une conscience de soi.

Tableau 3.5 : Indicateurs d'ajustement des modèles concernant la précarité de l'emploi

	D. L.	Déviante	Δ (%)	BIC	AIC
Chômage BIT :					
Indépendance cond.	37	1 601	1,32	1 153	1 527
Stabilité	36	26	0,16	-410	-46
Tendance linéaire	35	26	0,16	-398	-44
Splines	34	23	0,15	-389	-45
Halo du chômage :					
Indépendance cond.	37	2 921	2,52	2 472	2 847
Stabilité	36	52	0,29	-384	-20
Tendance linéaire	35	52	0,29	-373	-18
Splines	34	51	0,29	-361	-17
Emploi précaire ou halo du chômage :					
Indépendance cond.	30	2 207	2,84	1 847	2 147
Stabilité	29	25	0,26	-323	-33
Tendance linéaire	28	25	0,26	-311	-31
Splines	27	19	0,22	-305	-35

D. L. : degrés de liberté. Δ : indice de dissimilarité.

Plus encore que le niveau de l'endogamie relative, c'est sans doute la tendance temporelle qui constitue un test de l'apparition d'une nouvelle ligne de clivage de la structure sociale. En effet, le développement relativement récent de la précarité de l'emploi n'a pu entraîner immédiatement la constitution d'un groupe social : tout au plus, pourrait-on s'attendre à ce que la persistance durable d'un chômage de masse et d'un stock d'emploi précaires déploie progressivement ses effets sur certaines catégories sociales. À l'inverse, si

l'augmentation de la précarité ne s'est pas accompagnée du renforcement de la cohésion de la catégorie des précaires au cours du temps, il semble que l'on pourra répondre négativement à la question de savoir si un nouveau clivage est en train d'apparaître, remplaçant les anciennes distinctions en termes notamment de classes sociales.

C'est bien ce dernier scénario que nos données viennent soutenir : l'introduction d'une tendance temporelle n'améliore absolument pas la description des données. Le premier modèle testé, qui comprend une tendance temporelle linéaire, n'apporte presque rien au modèle de stabilité. Si les intervalles de confiance ne nous permettent pas d'exclure complètement la possibilité qu'une augmentation modérée soit cependant à l'œuvre, mais que notre échantillon soit trop faible pour la repérer, son ampleur maximale est limitée : les *odds ratios* partiraient toujours d'un niveau supérieur à 2 en 1975 et atteindraient difficilement le seuil de 4 en 2011³⁵. L'apparition d'une nouvelle ligne majeure de stratification sociale semble donc exclue du point de vue du choix du conjoint.

Supposer que l'endogamie ait suivi une tendance linéaire est cependant assez restrictif. La lecture de la figure 3.7 montre clairement que la précarité se développe par à-coups, et régresse parfois significativement au gré de la conjoncture économique. Afin de pouvoir identifier une tendance potentiellement irrégulière, nous testons une seconde spécification comprenant des splines cubiques restreintes³⁶, capables de représenter de manière flexible et cependant parcimonieuse des évolutions non nécessairement linéaires – par exemple une croissance globale suivant un rythme cyclique. Le résultat n'est pas plus concluant : plus on attribue de degrés de liberté à ces splines pour décrire de manière fine les évolutions temporelles, et moins les indicateurs BIC et AIC jugent le modèle performant.

Le meilleur modèle avec splines, qui n'ajoute qu'un degré de liberté au modèle avec tendance linéaire de manière à décrire une éventuelle évolution curvilinéaire, n'apporte cependant qu'une amélioration très limitée à la description des données. Ce modèle ne diminue jamais le BIC, et ne diminue l'AIC de manière significative (mais limitée) que concernant la définition la plus large de la précarité. Il décrit dans ce dernier cas une évolution légèrement concave, l'endogamie relative atteignant son minimum aux alentours

35 Le coefficient mesurant l'évolution temporelle de l'endogamie des précaires au sens le plus large est même négatif (et non significatif).

36 Nous présenterons plus en détail cette technique statistique lors de son utilisation plus intensive au chapitre 5 (Encadré 5.2, p. 308).

de 1997 – soit au moment où la proportion de couples composés de deux précaires atteint son sommet. Il apparaît en fait que l'endogamie relative du groupe des précaires varie en proportion inverse de l'extension de la catégorie : la corrélation entre la proportion d'individus en couple considérés comme précaires et l'*odds ratio* mesurant l'endogamie de ce groupe une année donnée est de -0,5. Cette corrélation ne se retrouve pas concernant les chômeurs. Tout indique donc que l'extension de la précarité, notamment liée à la situation économique, n'entraîne pas le développement sur le long terme d'une cohésion du groupe des précaires : les individus ne s'intègrent pas à un groupe, ils le rejoignent temporairement au gré du flux et reflux du stock d'emplois précaires.

Ces résultats sont tout à fait cohérents avec l'interprétation suivant laquelle « *l'effritement de la société salariale* » ne correspond pas à la constitution ou au renforcement d'un clivage entre inclus et exclus, mais au développement d'un continuum de situations plus ou moins proches du centre le plus intégré, pouvant aller jusqu'à la « *déstabilisation des stables* » (Castel, 1995, p. 715)³⁷. Selon la situation économique, la précarité de l'emploi touche un ensemble plus ou moins large de salariés, catégorie à géométrie variable qui ne constitue pas véritablement un groupe social.

6 Conclusion

Les analyses réalisées ici concernant la position des conjoints sur le marché du travail renforcent, tout en les nuanciant, les conclusions des chapitres précédents à propos d'autres dimensions de l'homogamie. Le développement de l'activité féminine explique la quasi-intégralité de la diminution des inégalités de salaires entre couples de près de 15 % qui a pris place entre 1990 et 2011. Contrairement à ce qui a eu lieu notamment aux États-Unis, ce mouvement ne s'est pas accompagné d'un renforcement de l'association entre salaires des conjoints : de ce fait, en France, l'homogamie a légèrement accentué la baisse des inégalités entre couples, particulièrement dans la moitié supérieure de la distribution. Ce résultat, relativement contre-intuitif, s'explique en partie par le fait que l'augmentation du taux d'emploi féminin a affecté de manière uniforme tous les milieux sociaux, mais aussi par un affaiblissement de l'association parmi les couples bi-actifs, qui signale ici encore une

37 La position de Robert Castel concernant cette question n'est pas restée constante selon les ouvrages, les plus récents ayant plutôt eu tendance à mettre l'accent sur le développement du « *précarariat* » (Castel, 2009).

diminution de l'homogamie. Rappelons cependant que nos données ne nous permettent pas de tenir compte des revenus autres que salariaux – dont notamment les revenus du patrimoine, qui apportent une contribution importante aux inégalités, en particulier au sommet de la distribution. Des études fondées sur des données plus riches seraient particulièrement utiles, même si elles ne remettraient vraisemblablement pas en cause nos principales conclusions.

Ainsi, nous n'observons pas l'apparition ou le renforcement de nouveaux clivages de la société française, qui viendraient contredire l'affaiblissement des dimensions classiques de la stratification sociale. Malgré une nette croissance de la proportion de couples formés de deux personnes en situation de précarité sur le marché du travail, l'endogamie relative de ce groupe reste faible au regard de celle des classes sociales ou des groupes de diplôme. Plus encore, aucune augmentation de cette endogamie n'est à relever. La constitution d'une classe sociale des précaires apparaît dès lors peu probable, comme pouvaient le laisser anticiper de nombreuses études existantes, qui pointaient la faible cohésion de cette catégorie très hétérogène et mouvante presque par définition.

L'évolution de l'hypergamie féminine est plus ambiguë. La contribution des femmes au total des salaires des couples, ainsi que la proportion des femmes qui gagnent plus que leur conjoint, ont augmenté de manière régulière et significative depuis trente ans, passant respectivement de 28 à 41 % et de 10 à 26 %. En outre, l'association entre salaires des conjoints parmi les couples bi-actifs se décrit mieux qu'avant par une comparaison directe entre salaire de l'homme et de la femme, plutôt que par une comparaison de leurs positions relatives par rapport à la moyenne de leur sexe.

En revanche, l'hypergamie féminine n'a pas diminué une fois tenu compte de l'évolution de la distribution des salaires des femmes (hypergamie relative) parmi les couples bi-actifs, signe que le choix du conjoint et les décisions d'activité parmi ces couples ne sont pas allées au-delà des évolutions de l'emploi et des salaires des femmes. On semble devoir conclure à l'existence de deux tendances contradictoires : d'une part, une augmentation de l'activité et des salaires des femmes, qui a évidemment impliqué une évolution des rôles de genre ; de l'autre, une résistance de la préférence pour la supériorité du salaire masculin au sein du couple, qui s'accommode de la première tendance mais n'en est absolument pas ébranlée. De fait, les études par entretiens montrent qu'il n'est pas rare

que les carrières féminines soient bridées par la crainte, partagée par les deux conjoints, que la femme gagne plus que l'homme (Testenoire, 2008). Ce paradoxe est à rapprocher de la lenteur des progrès de l'égalité entre les sexes dans certains domaines de la vie familiale, comme les tâches domestiques (Ricroch, 2012), dont le lien avec les écarts de salaire au sein du couple est avéré (Ponthieux et Schreiber, 2006).

Dans les prochains chapitres, nous systématiserons l'analyse de l'hypergamie féminine en nous intéressant aux autres caractéristiques des conjoints ainsi qu'en tenant compte des différentiels de célibat entre hommes et femmes. Nous verrons que le salaire est la dimension des différences genrées qui a le plus résisté, les autres présentant des évolutions bien plus importantes.

PARTIE II

MÉCANISMES DE L'HOMOGAMIE ET DE L'HYPERGAMIE EN FRANCE

CHAPITRE 4

LA DISPARITION DE L'HYPERGAMIE RELATIVE DE DIPLÔME, DE CLASSE ET D'ORIGINE SOCIALES : UNE CONVERGENCE DU CÉLIBAT ENTRE LES SEXES (1969-2011)

Nous avons mis en évidence dans la partie précédente un affaiblissement significatif – quoique avec des nuances –, au cours des quarante dernières années, de l'homogamie au sein des couples cohabitants, du point de vue de plusieurs critères. Un tel constat, se concentrant sur la composante symétrique entre hommes et femmes du choix du conjoint, serait cependant incomplet sans une étude systématique de l'évolution des différences genrées (de Singly, 1987a). En effet, comme nous l'avons défendu en introduction, le choix du conjoint ne révèle pas seulement la plus ou moins grande proximité entre groupes sociaux (homogamie) : à travers les asymétries entre hommes et femmes, il exprime aussi l'ordre de genre. La famille constituant une institution centrale de socialisation et de reproduction sociale, on peut soupçonner que la construction d'une position d'infériorité pour la femme au sein de celle-ci constitue l'un des facteurs essentiels de la domination masculine et de son maintien au fil des générations. Il semble donc intéressant de se demander si les évolutions majeures du statut des femmes qui ont pris place en France au cours du XX^e siècle ont remis en cause l'hypergamie féminine.

Nous avons déjà pu noter, dans nos analyses précédentes, des évolutions concernant les différences entre hommes et femmes. Ainsi, au chapitre 1, nous avons souligné que la massification scolaire a été plus forte du côté des femmes, entraînant leur supériorité par le diplôme au sein du couple à partir de l'an 2000. En revanche, les évolutions de la structure professionnelle ont été moins marquées : si les femmes occupent plus souvent des positions de cadres et professions intellectuelles qu'auparavant, la sur-représentation des hommes au sein de ces groupes persiste ; en outre, la ségrégation genrée au sein des classes populaires –

entre ouvriers et employées – s'est accentuée. Ces évolutions semblent être allées parfois au-delà de ce qu'imposait la transformation de la structure de la population (hypergamie relative) : nous avons ainsi pu noter au chapitre 2 une tendance à la diminution des asymétries entre hommes et femmes dans le choix du conjoint en termes de diplôme.

Finalement, nous avons identifié au chapitre 3 des évolutions majeures du point de vue des salaires des conjoints : plus souvent actives, les femmes ont aussi vu leur contribution au total des salaires des couples augmenter sensiblement depuis les années 1980. En revanche, la tendance à l'hypergamie relative parmi les couples bi-actifs, une fois contrôlé l'effet de l'augmentation de l'activité et des salaires féminins, n'a pas cédé d'un pouce.

Ces résultats, qui révèlent des changements parfois radicaux, d'autres fois plus ambigus, incitent à actualiser une conception datée des rapports entre genre, statut social et mise en couple qui semble persister, en décalage avec des évolutions sociales majeures déjà relativement anciennes. Ainsi, nous montrerons dans ce chapitre et le suivant que les évolutions de l'hypergamie de diplôme, de classe et d'origine sociales depuis 1969 ont été plus importantes que le discours sociologique ne l'admet généralement. Associé à d'autres évolutions majeures, comme le développement massif de l'activité féminine, ce mouvement semble de nature à contribuer à une modification des rapports de force au sein du couple, et donc à la remise en cause des rôles assignés aux hommes et aux femmes – même s'il est sans doute en lui-même insuffisant pour aboutir seul à un tel résultat.

Dans cette partie, nous étendrons notre étude à l'ensemble des individus, en couple ou célibataires, de manière à appréhender le fonctionnement du marché conjugal dans son ensemble. Cette modification du champ d'étude, qui nous permettra d'approfondir certains des mécanismes du choix du conjoint, se révèle essentielle pour l'étude de l'hypergamie féminine : en effet, l'une des différences fondamentales entre hommes et femmes du point de vue de la mise en couple tient à la distribution socialement très contrastée du célibat selon le sexe.

Ainsi, le phénomène de l'hypergamie sera ici divisé en trois composantes différentes : l'hypergamie féminine absolue, soit la proportion de couples dans lesquels l'homme occupe une position supérieure à celle de sa conjointe ; l'hypergamie féminine relative, soit la tendance à former ce type de couple au-delà des contraintes structurelles dues à la composition de la population ; le célibat asymétrique selon le sexe. Nous montrerons, dans

ce chapitre et le suivant, qu'hypergamie féminine et différences de célibat selon le sexe sont intimement liés : l'existence de taux de célibat différenciés selon le sexe et le statut social est une conséquence directe de l'asymétrie des préférences des hommes et des femmes sur le marché conjugal.

Nous commençons par présenter en détail le cadre d'interprétation classique de l'hypergamie, de sa signification, de ses conséquences et des mécanismes qui en sont à l'origine, puis avancerons des éléments concernant l'évolution de ce modèle fortement genré depuis les années 1960. Nous mettrons ensuite en évidence une inversion de l'hypergamie de diplôme et un affaiblissement très marqué de l'hypergamie de classe sociale, d'abord en termes absolus, puis en termes relatifs – soit en contrôlant l'évolution de la structure de la population à l'aide de modèles log-linéaires et log-multiplicatifs. Ces évolutions font écho à une certaine convergence entre hommes et femmes des taux de célibat par diplôme, classe et origine sociales. Une synthèse des évolutions de l'hypergamie et de la composante genrée du célibat réalisée à l'aide d'une technique peu utilisée, le modèle log-multiplicatif d'association anti-symétrique, nous permettra pour finir de faire ressortir la logique commune à ces deux phénomènes. Cette question sera approfondie au chapitre suivant à partir de données longitudinales.

1 Perspectives théoriques et état des connaissances

1.1 La conception classique de l'hypergamie féminine : un modèle fortement genré

L'hypergamie féminine ne peut se comprendre que dans le cadre d'une société où la domination masculine prévaut : elle est en effet à la fois une conséquence et un moyen de perpétuation de cette norme. Dans un modèle où les femmes se trouvent d'une manière générale en position inférieure aux hommes, il est essentiel pour la reproduction de l'ordre social que cet ordre soit d'abord respecté au sein de la famille, institution centrale d'intériorisation des rôles de genre (Goffman, 2002, sect. 6.2). Cette logique implique que les hommes prennent pour conjointes des femmes de statut social inférieur, et inversement, du fait à la fois de l'intériorisation de normes dominantes (Bozon, 1990a, p. 356-359), et de la volonté d'échapper aux sanctions sociales – explicites ou voilées – ou simplement aux

remarques négatives auxquelles sont exposés les couples qui dérogent à ces normes (Lefeuvre, 2008 ; Testenoire, 2008). La supériorité des hommes au sein du couple garantit aussi que dans la plupart des situations sociales soient spontanément répétés les rôles assignés aux membres de chaque sexe, consolidant ainsi les stéréotypes de genre contre une réalité qui peut aller à leur rencontre : c'est par exemple le cas en ce qui concerne l'assignation des tâches physiques aux hommes si l'on considère la tendance avérée à ce que l'homme soit plus grand (et donc, par convention, plus fort) que sa conjointe plus souvent que le hasard ne l'exigerait (Goffman, 2002, sect. 7 ; Herpin, 2003). L'hypergamie féminine permet aussi l'assignation persistante de la femme aux tâches domestiques (de Singly et Glaude, 1986), et justifie le sacrifice de sa carrière, moins prometteuse, à celle de l'homme (Nicole-Drancourt, 1989 ; de Singly, 1987b). À l'inverse, la supériorité du salaire de la femme amène à une gestion plus égalitaire de l'argent au sein du ménage (Testenoire, 2008).

Dans ce modèle parsonien (Parsons et Bales, 1955) où les hommes obtiennent généralement des revenus et un statut social supérieurs aux femmes sur le marché du travail, et où ces dernières sont supposées donner la priorité à la sphère familiale et à la représentation du couple envers l'extérieur, les comportements des individus respectent spontanément la norme d'hypergamie féminine : les femmes valorisent chez de potentiels conjoints les signes d'une (future) réussite sur le marché du travail, et les hommes valorisent chez de potentielles conjointes des qualités esthétiques, domestiques et relationnelles (Löwy, 2006, chap. 3). Ainsi, les hommes en couple dans les années 1980 en France déclaraient avoir apprécié, au moment de la mise en couple, que leurs conjointes soient « simples », « soigneuses », « jolies », « souriantes », « élégantes », « spontanées », « gaies » ; les femmes, que leurs conjoints soient « travailleurs », « affectueux », « calmes », « sécurisants » (Bozon, 1990b, p. 588 ; voir aussi South, 1991 ; Sprecher, Sullivan et Hatfield, 1994 pour les États-Unis).

Il n'est donc pas étonnant de constater que les hommes les plus diplômés, ou ceux occupant des professions prestigieuses, choisissent et soient choisis par des femmes dont la valeur matrimoniale ne consiste pas en des titres académiques, mais en des caractéristiques plus difficiles à mesurer par les enquêtes quantitatives¹. À l'appui de cette conception, on

1 On peut noter avec Oppenheimer (1995, p. 239) que la nature des capitaux spécifiques aux femmes reste généralement très floue dans les travaux portant sur la mise en couple, alors que les capitaux masculins sont très précisément définis.

sait par exemple (de Singly, 1982, 1987b) que les hommes diplômés disposent de revenus plus élevés lorsqu'ils sont mariés que lorsqu'ils sont célibataires, alors que c'est l'inverse chez les femmes ; ou encore que les ingénieurs de sexe masculin connaissent une carrière plus remarquable lorsqu'ils sont pères de nombreux enfants (Gadea et Marry, 2000).

Cette norme d'hypergamie a des conséquences importantes sur le célibat des groupes qui se trouvent défavorisés par leur position sur le marché conjugal : les femmes situées en haut de l'échelle des diplômes ou des professions, et les hommes situés en bas de celle-ci. Ainsi, parmi les générations nées avant guerre en France, le niveau de diplôme présentait des effets opposés selon le sexe, augmentant les chances de vivre en couple chez les hommes et les diminuant chez les femmes (Desplanques, 1987 ; Courgeau, 1987, p. 65 ; Robert-Bobée et Mazuy, 2005, p. 185 ; Winkler-Dworak et Toulemon, 2007, p. 294). Ce phénomène a aussi été observé de longue date du point de vue du statut socioprofessionnel des individus (Roussel, 1971, 1975 ; Villac, 1984). En particulier, dans une société française restée assez tardivement très rurale, le célibat des hommes agriculteurs, dû à l'exode des filles issues de ce groupe, a longtemps été la conséquence de l'hypergamie féminine la plus importante en termes d'effectifs concernés (Bourdieu, 1962 ; Jegouzo, 1972 ; Courgeau et Lelièvre, 1986).

Ce mécanisme genré favorise la sur-représentation des couples dans lesquels l'homme est le plus diplômé, par rapport à ce que l'on observerait si les individus prenaient un conjoint au hasard au sein de leur génération. En effet, les hommes occupant les positions les plus élevées choisissant moins souvent qu'on ne l'attendrait des femmes de leur niveau social, pour leur préférer des femmes de niveau (légèrement) inférieur, et les hommes situés immédiatement en dessous d'eux reproduisant ce comportement, un décalage apparaît et se propage, de proche en proche, le long de l'échelle sociale. Ce décalage condamne *in fine* les hommes situés en bas de la hiérarchie et les femmes situées à son sommet au célibat (Villac, 1984), étant entendu que les individus de ces deux groupes ne peuvent qu'exceptionnellement former une union ensemble du fait d'une distance sociale trop importante, redoublée par le fait que l'homme se trouverait dans une situation d'infériorité trop criante par rapport à sa conjointe.

Le célibat des femmes diplômées ou cadres et professions intellectuelles supérieures nous intéresse particulièrement ici : en effet, la norme de genre qui en est à l'origine va directement à l'encontre des logiques de reproduction sociale selon lesquelles les membres

des classes supérieures tendent à s'unir en priorité avec des individus de même niveau social. On peut avancer que dans ce modèle, c'est l'origine sociale qui compte, plus que le diplôme ou la profession, dans la définition du statut social d'une femme. Néanmoins, comment expliquer qu'une position plus élevée puisse avoir un effet négatif sur les chances de vie en couple d'une femme ? On peut envisager deux mécanismes expliquant ce phénomène, qui a été étudié principalement sous l'angle du diplôme.

Premièrement, suivant François de Singly (1987b, p. 176), on peut voir la source du célibat des diplômées dans « *un défaut d'articulation entre le calendrier de la formation de la dot scolaire et le calendrier de présentation sur le marché du mariage* ». En effet, la poursuite des études étant jugée incompatible avec la mise en couple, en particulier avant le développement de la contraception, les femmes diplômées n'étaient pas disponibles sur le marché conjugal jusqu'à un âge relativement avancé. Dans ces conditions, les hommes préféraient se mettre en couple avec des femmes peu diplômées mais ayant terminé leurs études – étant entendu que choisir une femme plus âgée ayant déjà décroché un diplôme élevé contrevenait à la norme d'hypergamie féminine en termes d'âge aussi bien que de capital culturel. Ce mécanisme laissait peu de chances aux femmes diplômées encore célibataires au-delà d'un certain âge de trouver un conjoint. Les professions les plus élevées étant occupées par les plus diplômées, ce mécanisme peut aussi expliquer les différences de célibat par classe sociale.

Un deuxième élément d'explication tient à la désirabilité d'une mise en couple, aussi bien du point de vue des intéressées que de leurs conjoints potentiels. D'un côté, les diplômées et les femmes exerçant les professions les plus qualifiées, ayant plus à perdre sur le marché du travail, étaient moins prêtes que les autres femmes à sacrifier leur carrière professionnelle à une vie familiale contraignante ou à une vie de couple inégalitaire (Cacouault, 1984 ; Kaufmann, 1999). Ces attentes relativement élevées étaient renforcées par l'indépendance financière que leur offrait leur activité professionnelle. De l'autre, aux yeux des hommes, un diplôme élevé ou un fort investissement professionnel pouvait être considéré comme un signal négatif indiquant un désintérêt à l'égard de la sphère familiale, un manque de féminité, et une indépendance contraire à l'idée qu'ils se faisaient du couple (Bihl et Pfefferkorn, 2002, p. 114 ; de Singly, 1987b, chap. 8).

Au total, dans ce modèle fortement genré, l'hypergamie féminine se décline en trois composantes que nous étudierons successivement. Tout d'abord, au sein des couples formés, l'homme possède le diplôme le plus élevé plus souvent que l'inverse. Ensuite, cette situation persiste lorsque les structures par diplôme des populations masculine et féminine sont contrôlées (hypergamie dite relative) : l'hypergamie est renforcée par un comportement qui va au-delà des contraintes de disponibilité des conjoints. Enfin, à côté de couples à hypergamie féminine majoritaire, le fonctionnement fortement genré du marché conjugal produit un nombre important d'exclus, à des positions inversées selon leur sexe : femmes dans les positions supérieures, hommes dans les positions inférieures.

1.2 L'évolution temporelle de l'hypergamie : une transformation majeure mais incomplète ?

Ce tableau très schématisé décrit les cohortes nées dans la première moitié du XX^e siècle, pour lesquelles la division sexuelle des rôles est fortement établie². La relative égalisation des statuts des hommes et des femmes réalisée au cours du XX^e siècle, visible notamment à l'élévation du niveau d'éducation des femmes, à leur entrée massive sur le marché du travail, au développement de la contraception, à la fin de la tutelle du mari sur sa femme, a profondément modifié ce modèle. Quels ont été les effets de cette évolution sur l'hypergamie ?

On peut tout d'abord s'attendre à un rapprochement des critères de choix du conjoint des hommes de ceux auparavant privilégiés par les femmes, un diplôme élevé et une carrière professionnelle n'étant plus considérés chez une femme comme un parcours déviant, voire comme le signe d'un manque de féminité. L'évolution des normes et des comportements et celle des stratégies de reproduction sociale convergent de ce point de vue. En effet, la massification scolaire a fait du diplôme un facteur essentiel de la stratification sociale dans tous les groupes sociaux (Beaud et Pialoux, 1999 ; Poullaouec, 2010). Étant donné que c'est la mère qui assure le plus souvent l'aide aux devoirs (Héran, 1994), son niveau d'éducation a un effet majeur – plus fort que celui du père – sur la réussite scolaire des enfants (Attias-

2 À l'encontre d'une lecture évolutionniste renvoyant l'hypergamie féminine à une pratique « traditionnelle » immémoriale, soulignons que l'hypergamie masculine était courante parmi les familles nobles de l'Occident chrétien des XI^e et XII^e siècles, le fort célibat des cadets engendrant une sur-représentation des femmes par rapport aux hommes à marier, et donc des possibilités d'ascension sociale par le mariage pour les hommes et leurs familles (Lett, 2000, chap. 6).

Donfut et Barnay, 2002 ; Place et Vincent, 2009). Sans supposer que les individus appliquent une rationalité implacable jusque dans leurs choix amoureux, on peut penser que cette situation nouvelle contribue à l'évolution de la perception des qualités d'une conjointe potentielle, en atténuant nettement l'image d'incompatibilité entre diplôme et vie familiale.

Par ailleurs, en termes purement économiques, si le diplôme et l'emploi des femmes pouvaient être perçus comme un signal négatif au sein des générations d'avant-guerre, un tel calcul a très clairement perdu en pertinence. Le développement du chômage, en augmentant l'incertitude attachée aux carrières masculines, rend particulièrement risquée la stratégie impliquée par le modèle de Gary Becker (1981), suivant laquelle l'homme aurait intérêt à se spécialiser dans l'activité sur le marché du travail, et la femme dans la sphère domestique, de manière à bénéficier de la croissance des rendements : l'emploi de la femme, même s'il est moins rémunérateur et fortement soumis à la précarité, peut s'avérer une garantie appréciable en cas de licenciement (Oppenheimer, 1994, 1995). De fait, les hommes mentionnent plus fréquemment qu'auparavant les questions financières comme critère de choix du conjoint (Buss *et al.*, 2001 pour les États-Unis). Cet argument prend tout son poids dans le contexte français, où les cohortes nées à partir des années 1960 ont fait face à un chômage élevé, alors que les carrières féminines, moins favorables que celles des hommes, sont cependant meilleures que dans nombre de pays occidentaux (Kempeneers et Lelièvre, 1993).

Mais l'évolution de l'hypergamie ne résulte pas exclusivement de celle des préférences et stratégies masculines : les préférences des femmes sur le marché conjugal peuvent elles aussi avoir changé, se rapprochant de celles des hommes (Oppenheimer, 1988, p. 584). Leur activité professionnelle leur permet de plus en plus de patienter avant de s'engager jusqu'à avoir rencontré un conjoint qui correspond à leurs attentes. Moins dépendantes de l'emploi de leur futur conjoint, elles peuvent valoriser d'autres qualités chez les hommes, ce qui se retrouve dans l'évolution des adjectifs qu'elles utilisent pour décrire leurs conjoints : moins souvent dans le registre du travail et de la protection, et plus dans celui de la sensibilité ou du physique (Bozon, 1990b, p. 592 ; Buss *et al.*, 2001).

Le diplôme est par ailleurs la dimension du statut social qui a connu les plus fortes transformations du point de vue des différences hommes-femmes. La massification scolaire

qui a pris place en France depuis la Seconde guerre mondiale a affecté les deux sexes, mais a été nettement plus marquée chez les femmes (Baudelot et Establet, 1992). L'élévation constante de leur niveau de diplôme par rapport à celui des hommes a nécessairement eu pour conséquence de diminuer la part des couples dans lesquels la femme est moins diplômée que son conjoint. Plus encore, les femmes étant désormais globalement plus diplômées que les hommes, l'hypergamie s'est inévitablement retournée en leur faveur.

Ainsi, nous pouvons nous attendre à observer un affaiblissement de l'hypergamie féminine selon ses différentes dimensions. Les femmes étant désormais globalement plus diplômées que les hommes, il était inévitable que l'hypergamie de diplôme se soit inversée en termes absolus – ce que nous avons déjà observé au chapitre 1. Cette évolution se distingue des mouvements, plus limités et localisés, de brouillage de la division sexuelle du travail, et constitue un cas d'« inversion du genre » (Kergoat, Guichard-Claudic et Vilbrod, 2008) à grande échelle. L'importance des transformations à l'œuvre exclut qu'elles aient pu prendre place sans une modification profonde des normes de genre, et qu'elles n'aient eu en retour un effet plus large sur ces dernières (Baudelot et Establet, 1992, Avant-propos). Jean-Pierre Terrail (1995) voit dans la plus grande réussite scolaire des filles le signe d'une mobilisation visant à acquérir une indépendance supérieure à ce qu'elles observent parmi les générations de leurs mères. Participant au mouvement plus large d'émancipation des femmes et porté par l'augmentation de leur activité professionnelle, l'allongement des carrières scolaires féminines n'a rien d'un épiphénomène. On pourrait donc en déduire que l'hypergamie féminine en termes relatifs (c'est-à-dire une fois contrôlée l'évolution de la structure de la population) devrait s'être affaiblie, voire avoir disparu, en termes de diplôme comme en termes de profession ou d'origine sociale, et que les différences de célibat entre hommes et femmes se soient atténuées.

Néanmoins, tout n'est pas si clair. En effet, si la participation des femmes au marché du travail a explosé entre les premières et les dernières cohortes que nous étudions, et si la nature même de l'emploi des femmes a changé depuis la période d'après-guerre, l'égalité en termes de carrière et de salaire entre hommes et femmes est loin d'être atteinte, et les femmes gagnent en moyenne toujours nettement moins que leur conjoint (*cf.* chapitre précédent). On a vu que l'hypergamie relative de salaire n'a pas diminué parmi les couples

bi-actifs. De ce fait, on pourrait s'attendre à observer un affaiblissement de l'hypergamie féminine, mais un affaiblissement qui ne serait que partiel, une fois contrôlée la structure de la population. Cette hypothèse revient à avancer que les normes sociales opposeraient une certaine résistance, ou du moins une inertie, aux évolutions à l'œuvre.

On peut en outre penser que cette inertie diffère selon les dimensions considérées. Il serait ainsi nécessaire de distinguer hypergamie de diplôme d'un côté, et hypergamie de classe sociale et de salaire de l'autre, ces deux derniers aspects constituant le cœur des rôles masculins, quand le premier n'en serait qu'un élément secondaire. L'évolution de l'hypergamie refléterait alors celle des inégalités de genre dans chaque domaine.

Au contraire de cette hypothèse, fondant leur argument sur la persistance des normes de genre, Hans-Peter Blossfeld et Andreas Timm (2003, p. 10) vont jusqu'à avancer que la solidité du modèle du *male breadwinner* devrait conduire au maintien de l'hypergamie féminine en termes de diplôme malgré la hausse du niveau d'éducation des femmes, condamnant les femmes les plus diplômées et les hommes les moins diplômés au célibat. Néanmoins, cette argumentation repose sur l'hypothèse qu'il existe réellement une norme d'inactivité féminine et qu'elle est assez profondément ancrée pour résister à des évolutions structurelles de grande ampleur. On peut douter de la pertinence de cette lecture appliquée à la France, les auteurs l'ayant dérivée d'une étude pilote concernant l'Allemagne de l'Ouest, où les inégalités entre hommes et femmes sont notoirement plus marquées aussi bien du point de vue de l'éducation que de la carrière professionnelle (Marry *et al.*, 1998)³.

De fait, ainsi que nous l'avons annoncé plus haut, le scénario d'une persistance de la norme d'hypergamie est très clairement infirmé par l'inversion de l'hypergamie de diplôme et la quasi-disparition de l'hypergamie de classe sociale que nous observons. Cette norme semble ainsi bien moins prégnante qu'on ne pouvait le supposer, ce que confirment les travaux existants concernant de nombreux pays.

3 Dans les analyses de D. Monaghan (2014, p. 41, voir ci-dessous), les trois pays développés où l'hypergamie de diplôme est la plus marquée sont de loin la Suisse, l'Autriche et l'Allemagne. Ces résultats sont conformes à nos observations (chapitre 7).

1.3 Les travaux existants : baisse et inversion de l'hypergamie de diplôme

Comme pour les études portant sur l'homogamie, c'est le diplôme qui a le plus retenu l'attention des chercheurs, les autres dimensions n'ayant presque pas été étudiées⁴. La plupart des travaux récents s'étant attachés à décrire les évolutions de l'hypergamie de diplôme ont observé une diminution, et souvent même une inversion de celle-ci. Dans une analyse comparative très large incluant 56 pays de tous continents et niveaux de développement, Albert Esteve, Joan García-Román et Iñaki Permanyer (2012) ont observé une tendance temporelle très nette. Parmi les 18 pays pour lesquels les auteurs disposent de données portant sur le début des années 1970, l'hypergamie féminine était toujours plus fréquente que l'hypogamie chez les jeunes couples cohabitants. Au contraire, au début des années 2000, 26 des 51 pays inclus dans l'échantillon à cette date se trouvent dans la situation inverse. Les sociétés concernées sont très diverses, mais parmi les pays occidentaux étudiés on peut citer, avec la France, l'Espagne, les États-Unis, la Grèce, la Hongrie, l'Italie, le Portugal et la Slovaquie. Dans un document de travail plus détaillé (Esteve, García-Román et Permanyer, 2011), les mêmes auteurs ont étudié l'hypergamie relative, soit la tendance à l'hypergamie observée une fois contrôlée la distribution par diplôme des populations des hommes et des femmes. Il apparaît qu'aucune divergence entre hypergamie brute et hypergamie relative ne peut être observée : l'hypergamie féminine se trouve ainsi renforcée par le choix du conjoint dans les sociétés où les femmes sont globalement moins diplômées que les hommes, alors que c'est l'hypogamie qui l'est dans celles où elles sont plus diplômées. En d'autres termes, les comportements des individus semblent toujours prolonger et renforcer les contraintes que la structure sociale fait peser sur eux. Ce résultat semble remettre en question l'existence d'une norme d'hypergamie : dans aucun pays, une telle norme ne semble avoir résisté à l'évolution de la structure par diplôme en faveur des femmes.

Ces évolutions se trouvent confirmées par plusieurs travaux portant eux aussi sur l'hypergamie d'éducation. Alors que les études avaient jusque-là conclu à l'existence d'une tendance uniforme à l'hypergamie féminine (Smits, Ultee et Lammers, 1998, p. 274), Henryk Domański et Dariusz Przybysz (2007) ont pour la première fois relevé, dans une étude

4 Nous n'avons connaissance d'aucune étude portant sur un autre pays que la France concernant les évolutions de l'hypergamie de classe ou d'origine sociales.

portant sur 22 pays européens, que l'hypergamie féminine était loin d'être la norme parmi les couples mariés. Ainsi, dans 7 pays, une tendance statistiquement significative à l'hypogamie féminine en termes relatifs était relevée. Plus récemment, David Monaghan (2014) a observé que l'hypogamie relative des femmes est plus forte que leur hypergamie relative parmi les jeunes couples cohabitants dans 13 des 26 pays développés étudiés entre 1980 et 2010. Plusieurs travaux portant sur un pays en particulier ont relevé les mêmes tendances. C'est le cas des États-Unis, où Elaina Rose (2004) note une quasi-disparition de l'hypergamie de diplôme, Zhenchao Qian (1998) puis Christine Schwartz et Robert Mare (2005) ayant même mesuré une inversion, en termes absolus et relatifs. Le même mouvement a été observé par Albert Esteve et Clara Cortina (2006) en Espagne, ainsi que par Elina Mäenpää et Marika Jalovaara (2014) en Finlande. Enfin, ce phénomène a été relevé en Amérique du Sud par plusieurs travaux (Esteve et McCaa, 2007 ; López-Ruiz, Esteve et Cabré, 2009 ; Rodríguez, 2014).

Finalement, concernant la France, Yvonne Guichard-Claudic, Armelle Testenoire et Danièle Trancart (2009) ont mesuré en utilisant l'enquête Familles et employeurs (Ined, 2004-2005) une hypogamie de diplôme près de deux fois plus fréquente chez les femmes âgées de 25 à 30 ans que parmi celles âgées de 40 à 50 ans. Parmi ces générations nées entre 1955 et 1980, l'hypergamie féminine en termes de diplôme est toujours moins fréquente que l'hypogamie – mais c'est l'inverse en termes de catégorie socioprofessionnelle. Ces résultats rejoignent ceux déjà mis en évidence au chapitre 1, selon lesquels l'hypergamie féminine de diplôme est devenue moins fréquente que son contraire aux alentours de l'an 2000.

Les mêmes auteurs ont aussi réalisé l'un des rares études concernant l'évolution de l'hypergamie socioprofessionnelle dont nous ayons connaissance. Elles ont relevé la persistance de l'hypergamie féminine, et affirment – malheureusement sans fournir de détails – que cette tendance n'a pas évolué au fil des cohortes. Néanmoins, la critique déjà formulée au chapitre 1 à l'égard des études fondées sur une seule enquête transversale s'applique ici : les couples âgés ne peuvent être considérés comme représentatifs d'un état antérieur de la société, en particulier en ce qui concerne la profession, qui n'est pas stable au cours du temps. Or, une étude plus ancienne de Louis-André Vallet (1986), fondée sur les recensements de la population de 1962, 1968, 1975 et 1982, concluait à un léger affaiblissement de l'hypergamie féminine à la fois en termes absolus, mais aussi au-delà de

ce que l'évolution de la structure de la population exigeait. Nous restons donc prudents concernant l'évolution de l'hypergamie de classe sociale que nous devrions nous attendre à observer pour les trente dernières années.

Au total, contrairement à ce que les éléments théoriques présentés plus haut auraient laissé prévoir, dans de nombreuses sociétés, la norme d'hypergamie féminine, approchée par l'hypergamie relative, semble s'être inversée, du moins du point de vue du diplôme. Malheureusement, la question, indissociable de l'hypergamie, de l'évolution des taux de célibat selon le sexe et le diplôme n'a que plus rarement été étudiée. On peut cependant noter qu'aux États-Unis, E. Rose (2004) a observé que l'inversion de l'hypergamie s'est accompagnée d'une disparition du surcroît de célibat des plus diplômées (dont Goldstein et Kenney, 2001 notaient même l'inversion), mais aussi une augmentation du célibat des hommes les moins diplômés. Outre la question de l'évolution de l'hypergamie, il semble ainsi important d'évaluer si les chances de vie en couple ont suivi le même mouvement en France, et si un mouvement similaire s'observe concernant la classe et l'origine sociales.

2 L'hypergamie absolue : entre affaiblissement et inversion

Nous commençons par mesurer la prévalence de l'hypergamie de diplôme, de classe et d'origine sociales parmi les couples cohabitants au cours des quarante dernières années. Les évolutions majeures observées dans cette section seront approfondies dans la section suivante par l'étude de l'hypergamie relative.

2.1 La définition de l'hypergamie : hiérarchiser les catégories

L'identification de configurations de couple à hypergamie ou à hypogamie féminine exige de se doter d'un ordre hiérarchique des caractéristiques des conjoints. La définition de l'hypergamie de diplôme est immédiate, étant donné que le diplôme est une variable ordonnée⁵. Celle de l'hypergamie de classe sociale et de classe sociale d'origine est plus délicate à établir, étant donné que la classification EGP, qui comprend des catégories d'indépendants, n'est pas intégralement construite suivant une échelle de statut. Nous

5 Nous ne tenons pas ici compte des différences entre filières d'études, dont l'effet sur les inégalités professionnelles entre hommes et femmes est majeur (Duru-Bellat, 2004 ; Couppié et Epiphane, 2006), mais qui ne peuvent complètement compenser l'effet de la différenciation verticale des diplômes.

définissons donc une classification des configurations de couple selon qu'elles apparaissent comme endogames, à hypergamie féminine, à hypogamie féminine, ou non ordonnées. Afin de garantir la robustesse des résultats, tous les cas dans lesquels la hiérarchie entre classes des conjoints n'est pas absolument claire ont été classés dans ce dernier groupe.

Nous nous sommes inspirés pour l'établissement de ce classement de l'« ordre de dominance » utilisé par Robert Erikson et John Goldthorpe (1992, p. 266) afin d'identifier la classe sociale d'un ménage à partir de celles des deux conjoints, ainsi que de la matrice de mobilité socioprofessionnelle construite par Camille Peugny (2007) pour mesurer la mobilité sociale intergénérationnelle descendante. Nous avons cependant dû apporter quelques ajustements. En effet, l'approche des deux premiers auteurs ne vise pas à retenir la classe sociale la plus élevée des deux, mais celle qui définit le plus profondément la position sociale du ménage. Pour cela, les classes d'indépendants sont toujours considérées comme dominantes par rapport à celles d'employés qualifiés (IIIa) ou de techniciens et contremaîtres (V), ce qui ne paraît pas le plus approprié dans notre cas.

La solution retenue par le second auteur est quant à elle fondamentalement intergénérationnelle, et elle vise à tenir compte de l'évolution du statut social des professions au cours du temps. Pour cela, la matrice n'est pas parfaitement symétrique, un enfant d'employé ou ouvrier qualifié étant devenu artisan ou commerçant étant considéré comme immobile, quand la trajectoire inverse est considérée comme descendante. Même si les hommes et les femmes n'occupent généralement pas les mêmes postes au sein d'une même classe sociale, il ne nous semble pas justifié d'introduire ce type de distinction qui, faute de s'appuyer sur des données objectives, s'exposerait au risque de l'arbitraire. La matrice que nous retenons est donc parfaitement symétrique⁶, et nous faisons l'hypothèse que la ségrégation genrée entre professions au sein d'une classe sociale ne remet pas en cause leur hiérarchie.

6 Ou plus rigoureusement, symétrique pour les configurations non ordonnées (configurations endogames et non classées), et anti-symétrique pour les autres (une configuration à hypergamie féminine étant le symétrique par rapport à la diagonale d'une configuration à hypogamie féminine).

Tableau 4.1 : Matrice d'hypergamie de classe sociale

Hommes	Femmes							
	I	II	V	IIIa & VI	IIIb & VIIab	IVa	IVb	IVc
I	H=F							
II	H<F	H=F						
V	H<F	H<F	H=F					
IIIa & VI	H<F	H<F	H<F	H=F				
IIIb & VIIab	H<F	H<F	H<F	H<F	H=F			
IVa	-	H>F	H>F	H>F	H>F	H=F		
IVb	H<F	-	-	-	H>F	H<F	H=F	
IVc	H<F	-	-	-	H>F	H<F	-	H=F

« H=F » indique que les classes sociales des deux conjoints sont considérées comme approximativement de même statut. « H<F » indique que le statut social de la femme est supérieur à celui de son conjoint (hypogamie féminine), alors que H>F indique la situation inverse (hypergamie féminine). « - » indique l'absence de classement.

La partie supérieure de la table est l'inverse de sa partie inférieure.

Cf. Encadré 1.1, p. 72 pour l'intitulé des classes EGP.

La matrice retenue (Tableau 4.1) suit quelques principes simples. Le classement des catégories salariées ne pose pas de difficulté particulière. Du fait de la très forte ségrégation genrée au sein des classes populaires soulignée au chapitre 1, nous considérons ouvriers et employés qualifiés d'un côté, ouvriers et employés non qualifiés de l'autre, comme étant de même statut. Ce regroupement, aussi retenu par C. Peugny (2007), apparaît pleinement justifié par des travaux déjà cités (Burnod et Chenu, 2001 ; Amossé et Chardon, 2006). Il va néanmoins à l'encontre d'un classement fondé sur le choix du conjoint : sur l'échelle mesurée au chapitre 2 (Figure 2.6, p. 167), les employés qualifiés (IIIa) se détachent nettement du reste des catégories populaires. Soulignons donc que la hiérarchie des classes sociales établie ici se fonde sur des critères externes à l'homogamie ; ce dernier principe de classement sera mobilisé dans la section suivante.

C'est le classement des configurations comprenant un indépendant qui est le plus délicat. Les indépendants avec salariés (IVa) ont été considérés comme de statut supérieur à toutes les autres catégories, à l'exception des cadres et professions intellectuelles

supérieures (I), avec lesquels ils forment une paire non ordonnée : en effet, une partie importante de ce groupe est constituée de chefs d'entreprises de taille relativement modeste – les gros indépendants étant déjà intégrés dans la classe I⁷. De leur côté, les indépendants sans salariés (IVb) et les agriculteurs (IVc) ont été traités de la même manière. Étant donné la grande hétérogénéité de ces groupes, nous avons renoncé à les considérer comme hiérarchiquement ordonnés par rapport aux cadres et professions intellectuelles de rang inférieur (II), aux techniciens et contremaîtres (V) ainsi qu'aux ouvriers et employés qualifiés (VI, IIIa) ; ces deux classes ne sont pas non plus ordonnées entre elles. En revanche, elles sont considérées comme de rang inférieur aux cadres et professions intellectuelles supérieures (I) et aux indépendants avec salariés (IVa), ainsi que de rang supérieur aux ouvriers et employés non qualifiés (VIIab, IIIb).

Cette hiérarchie est appliquée de manière identique aux classes d'origine des conjoints, la classe des indépendants avec ou sans salariés (IVab) étant considérée comme celle des indépendants sans salariés (IVb), et le croisement avec la catégorie I comme non ordonné.

Pour finir, notons que différentes variantes de ce classement fournissent des résultats similaires du point de vue des évolutions temporelles.

2.2 L'inversion de l'hypergamie absolue de diplôme

Les évolutions de l'hypergamie absolue de diplôme, variable plus immédiatement ordonnée, ont déjà été commentées au chapitre 1 (Figure 1.4, p. 96). Rappelons simplement que nous avons relevé une augmentation nette de la proportion des couples dans lesquels la femme est la plus diplômée. Au contraire, la proportion des couples dans lesquels l'homme est le plus diplômé a légèrement diminué depuis les années 1980, alors qu'elle augmentait jusqu'à cette date. Depuis environ l'an 2000, il est plus fréquent que les femmes soient plus diplômées que leur conjoint que l'inverse. Ces évolutions s'expliquent au moins en grande partie par l'augmentation du niveau de diplôme des femmes par rapport aux hommes, comme l'indique la comparaison des tendances observées avec celles qui seraient attendues en situation de mise en couple au hasard. Il n'est donc pas évident qu'une tendance à l'hypergamie féminine selon le diplôme existe au-delà des contraintes structurelles.

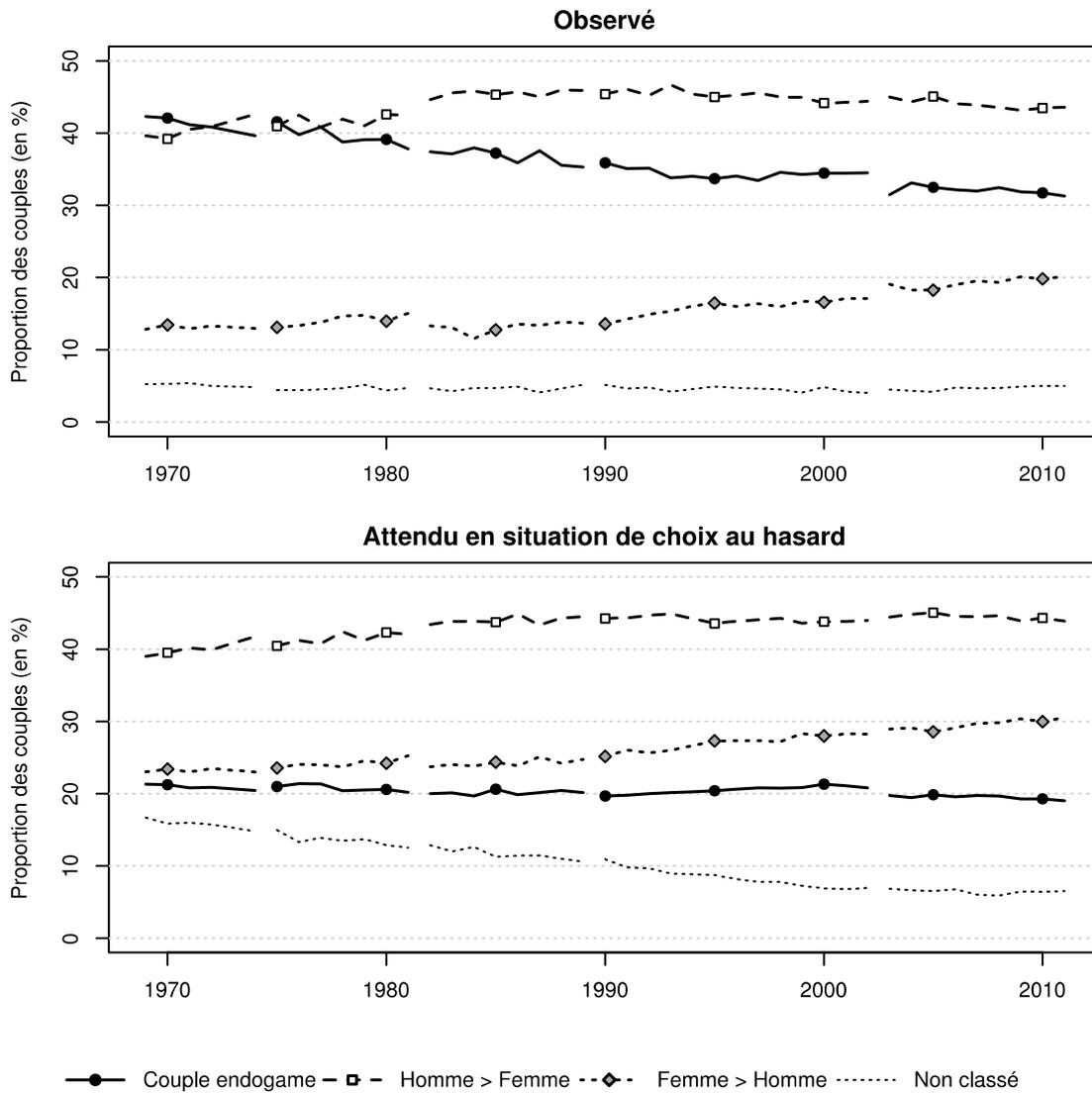
7 Du moins dans les séries qui le permettent, cf. Annexe A, p. 517.

2.3 Le développement de l'hypogamie féminine de classe sociale

La matrice d'hypergamie définie plus haut nous permet maintenant de réaliser la même analyse concernant la classe sociale des conjoints (Figure 4.1, graphique du haut). Outre la diminution du taux d'endogamie déjà relevée au chapitre 1, on observe une certaine stabilité de la proportion de couples dans lesquels l'homme appartient à une classe sociale supérieure à celle de sa conjointe (hypergamie féminine), quoique avec une légère croissance jusqu'en 1990 (avec un sommet à 45 %) suivie d'une légère décroissance⁸. L'évolution principale provient donc de la hausse de la proportion de couples dans lesquels la femme a un statut supérieur à son conjoint (hypogamie féminine), de 14 % en 1990 à 20 % en 2011. Ce mouvement, assez tardif, provient presque exclusivement du développement des couples dans lesquels la femme appartient aux cadres et professions intellectuelles supérieures (I et II), qui passent de 21 % à 35 % de l'ensemble des couples entre 1990 et 2011 ; les autres classes féminines apportent une contribution négligeable, voire négative, à cette évolution. Une analyse portant sur les seuls couples bi-actifs (non reprise ici) fournit des résultats très proches, avec pour seule différence une plus grande proportion de couples endogames et une moindre représentation des couples à hypergamie féminine, aboutissant à une légèrement plus grande égalité entre conjoints.

8 La rupture de série en 1982 est due au codage de la distinction entre employés qualifiés et non qualifiés, question déjà abordée au chapitre 1.

Figure 4.1 : Évolution des taux d'endogamie, d'hypergamie et d'hypogamie de classe sociale observés et attendus en situation de mise en couple aléatoire



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 30 à 59 ans et les deux conjoints ont déjà travaillé

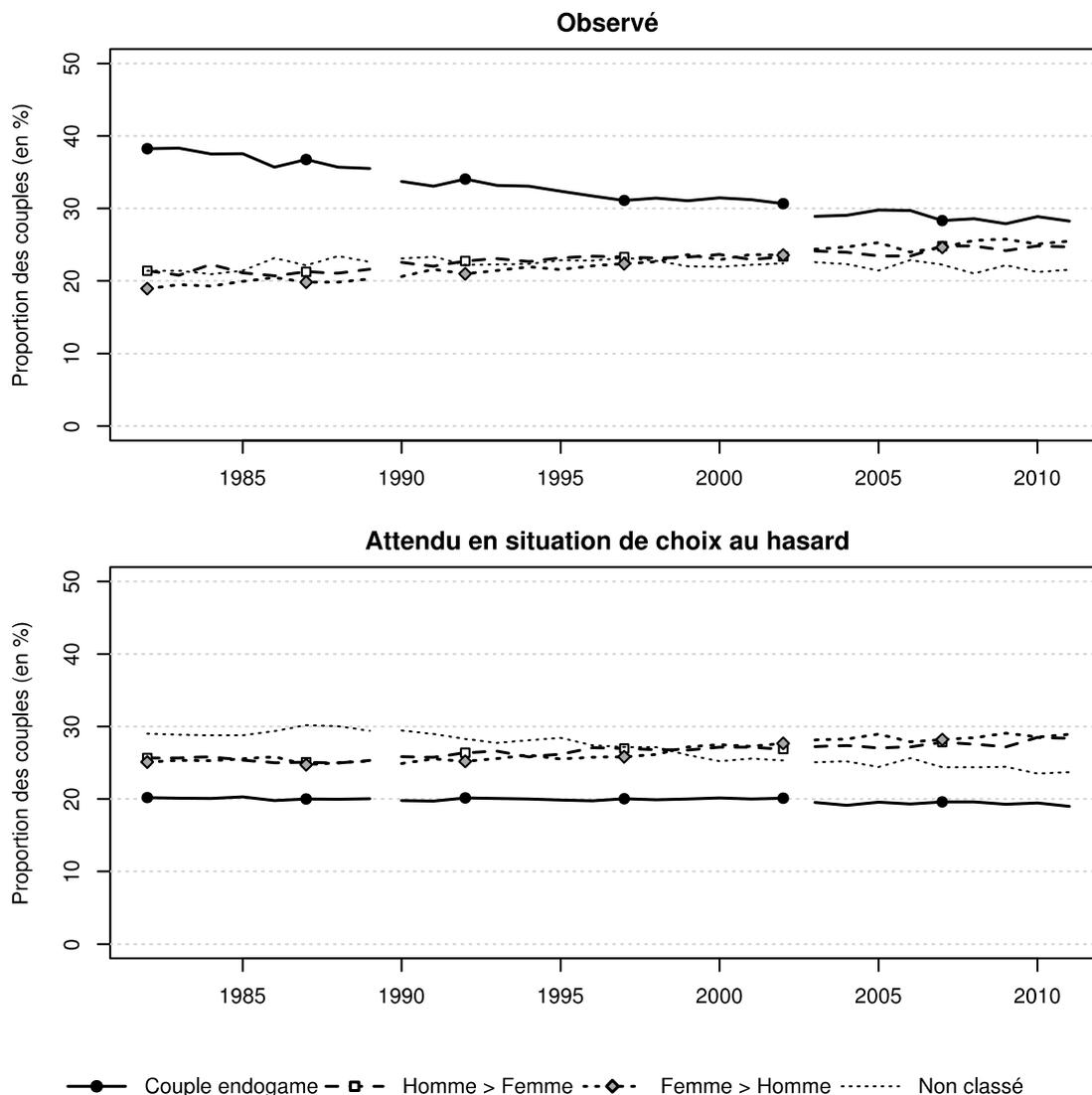
Ces tendances correspondent globalement à celles qui seraient attendues en situation de choix aléatoire du conjoint (Figure 4.1, graphique du bas). Même si on observe une proportion de couples à hypogamie féminine sensiblement inférieure à celle qui serait attendue en situation de mise en couple au hasard, ce sont avant tout la forte sur-représentation des couples endogames et la sous-représentation des couples non classés – due à l'association négative entre indépendants et salariés – qui ressortent. Ainsi, l'existence d'une norme d'hypergamie féminine du point de vue de la classe sociale n'apparaît pas aussi frappante que du point de vue du diplôme : c'est avant tout la position des hommes et des femmes sur le marché du travail qui semble à l'origine de la supériorité professionnelle des hommes dans le couple.

2.4 L'absence d'hypergamie selon la classe sociale d'origine

Cette conclusion est renforcée par l'étude de l'hypergamie d'origine sociale. Les classes sociales d'origine des hommes et des femmes étant par définition égales, l'hypergamie et l'hypogamie féminines seraient également fréquentes en situation de mise en couple au hasard. Toute différence entre hypergamie et hypogamie féminines dans les données observées serait donc le reflet des mécanismes genrés de choix du conjoint, et plus particulièrement du célibat différencié entre hommes et femmes selon leur origine sociale.

Mais aucun phénomène de ce type n'apparaît (Figure 4.2). Les couples dans lesquels la femme est d'une origine supérieure à celle de son conjoint sont pour toutes les années aussi fréquents que ceux dans lesquels l'inverse est vrai, aussi bien en termes de proportions attendues en situation de choix aléatoire du conjoint que de proportions observées. Ainsi, lorsque les contraintes structurelles ne l'imposent pas, il ne semble pas que les couples se forment très clairement suivant un modèle dans lequel l'homme doit avoir un statut supérieur à celui de sa conjointe. Afin de vérifier cette hypothèse, nous recourons dans la section suivante à des modèles log-linéaires capables de contrôler l'effet de la structure de la population.

Figure 4.2 : Évolution des taux d'endogamie, d'hypergamie et d'hypogamie de classe sociale d'origine observés et attendus en situation de mise en couple aléatoire



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 30 à 59 ans

3 La fin de l'hypergamie relative

Comme l'homogamie relative, l'hypergamie relative se définit en termes d'*odds ratios*. Mais alors que la première se mesure à l'écart de ces *odds ratios* à l'unité, qui correspond à la situation d'indépendance, l'hypergamie se mesure au rapport entre les *odds ratios* généralisés correspondant aux configurations à hypergamie féminine et leurs symétriques

par rapport à la diagonale, qui correspondent aux configurations à hypogamie féminine⁹. L'hypergamie relative désigne donc la sur-représentation des configurations dans lesquelles l'homme occupe une position supérieure à celle de sa conjointe par rapport à celles dans lesquelles l'inverse est vrai, une fois contrôlée la structure de la population.

L'hypergamie correspond donc exclusivement à la composante asymétrique de l'association : elle se définit comme écart à une association complètement symétrique, dans laquelle les positions des hommes et des femmes seraient interchangeable. Afin de la modéliser, nous prenons donc pour référence des modèles utilisés au chapitre 2 (diagonale exclue), mais en contraignant l'association à être parfaitement symétrique. Nous introduisons ensuite un paramètre mesurant la sur-représentation des configurations à hypergamie féminine. La diagonale étant exclue, ce paramètre oppose ces configurations à celles à hypogamie féminine. La valeur de ce paramètre est représentée sous la forme d'un *odds ratio* généralisé anti-symétrique (Encadré 4.1), qui est plus directement relié à la définition de l'hypergamie relative fournie au paragraphe précédent.

De manière à assurer l'indépendance des marges du modèle, il est important que la partie symétrique de l'association soit décrite correctement, même si elle ne présente pas en elle-même d'intérêt pour notre analyse. Nous prenons donc comme base le modèle à forme de régression, qui était apparu au chapitre 2 (Tableau 2.2, p. 156) comme le meilleur modèle, mais en remplaçant l'interaction complète entre lignes et colonnes par une association symétrique. Les équations des modèles sont détaillées dans l'encadré 4.1.

Encadré 4.1 : Équations des modèles

Tous les modèles prennent pour base le modèle de quasi-symétrie (Causinus, 1965). En notant m_{hft} les effectifs prédits par le modèle pour la cellule à l'intersection de la ligne h (catégorie de l'homme), de la colonne f (catégorie de la femme) et de la couche t (année d'enquête), dans une table de dimensions $H \times F \times T$, ce modèle s'écrit :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + s_{hf}^{HF}$$

avec $s_{hf}^{HF} = s_{fh}^{HF}$ les paramètres d'association symétriques.

⁹ Les configurations endogames, situées sur la diagonale de la table, sont leur propre symétrique et n'interviennent pas dans la mesure de l'hypergamie. De même, les configurations non classées, dont les symétriques ne sont pas non plus classées, sont ignorées.

Le modèle de quasi-symétrie à forme de régression, qui combine une évolution temporelle à forme de régression (Goodman et Hout, 1998) et une interaction entre lignes et colonnes symétrique, étend le modèle de quasi-symétrie. Il s'écrit :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + s_{hf}^{HF} + \psi_t s_{hf}^{HFT}$$

avec $s_{hf}^{HF} = s_{fh}^{HF}$ les paramètres d'association symétriques décrivant l'état de départ, $s_{hf}^{HFT} = s_{fh}^{HFT}$ ceux décrivant l'état d'arrivée, et ψ_t les coefficients annuels indiquant la position dans la transition entre les états de départ et d'arrivée.

Ce dernier modèle postule une association parfaitement symétrique. Plus généralement, l'association entre lignes et colonnes une année donnée peut toujours s'écrire $\lambda_{hft}^{HFT} = s_{hft}^{HFT} + a_{hft}^{HFT}$ avec $s_{hft}^{HFT} = (\lambda_{hft}^{HFT} + \lambda_{fht}^{HFT})/2$ la composante symétrique de l'association déjà décrite, et $a_{hft}^{HFT} = (\lambda_{hft}^{HFT} - \lambda_{fht}^{HFT})/2$ sa composante anti-symétrique. Les modèles suivants ajoutent au précédent des paramètres permettant de décrire de manière parcimonieuse cette seconde composante, qui correspond à l'hypergamie.

Le modèle de stabilité de l'hypergamie, qui complète le précédent, s'écrit :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + s_{hf}^{HF} + \psi_t s_{hf}^{HFT} + \delta^{HF} \mathbb{1}_{p_h > p_f} + \delta'^{HF} \mathbb{1}_{p_h \parallel p_f}$$

avec $\mathbb{1}_{p_h > p_f}$ indiquant que la femme occupe une position inférieure à celle de son conjoint et $\mathbb{1}_{p_h \parallel p_f}$ indiquant que les positions des conjoints ne peuvent être ordonnées. δ^{HF} mesure ainsi la sur- ou sous-représentation des configurations à hypergamie féminine par rapport à celles à hypogamie féminine, alors que δ'^{HF} mesure celle des configurations qui n'ont pas été classées. Les cellules de la diagonale étant exclues, les couples à hypogamie féminine (sans les couples endogames) constituent la référence. Sous ce modèle, l'association anti-symétrique s'écrit simplement $a_{hft}^{HFT} = \delta^{HF}/2$ pour les configurations ordonnées et $a_{hft}^{HFT} = \delta'^{HF}/2$ pour les configurations non ordonnées.

Le modèle de variation temporelle linéaire de l'hypergamie s'écrit :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + s_{hf}^{HF} + \psi_t s_{hf}^{HFT} \\ + (\delta + t\delta^T) \mathbb{1}_{p_h > p_f} + (\delta' + t\delta'^T) \mathbb{1}_{p_h \parallel p_f}$$

avec δ et δ' les paramètres mesurant la sur- ou sous-représentation des configurations respectivement à hypergamie féminine et non classées au début de la période, et δ^T et δ'^T ceux mesurant l'évolution d'une année sur l'autre. Sous ce modèle, on voit que l'association anti-symétrique s'écrit $a_{hft}^{HFT} = (\delta + t\delta^T)/2$ pour les configurations ordonnées et $a_{hft}^{HFT} = (\delta' + t\delta'^T)/2$ pour les configurations non ordonnées.

Enfin, le modèle de variation temporelle libre de l'hypergamie s'écrit :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + s_{hf}^{HF} + \psi_t s_{hf}^{HFT} + \delta_t^{\mathcal{T}} \mathbb{1}_{p_h > p_f} + \delta_t'^{\mathcal{T}} \mathbb{1}_{p_h \parallel p_f}$$

avec $\delta_t^{\mathcal{T}}$ et $\delta_t'^{\mathcal{T}}$ les paramètres mesurant chaque année t la sur- ou sous-représentation des configurations respectivement à hypergamie féminine et non classées. Sous ce modèle, on voit que l'association anti-symétrique s'écrit $a_{hft}^{HFT} = \delta_t^{\mathcal{T}}/2$ pour les configurations ordonnées et $a_{hft}^{HFT} = \delta_t'^{\mathcal{T}}/2$ pour les configurations non ordonnées.

La signification de ces paramètres peut se comprendre en repartant des *odds ratios* généralisés. En effet, on peut définir l'*odds ratio* généralisé symétrique $GOR_{hf(t)}^S$ comme la moyenne géométrique de l'*odds ratio* généralisé correspondant à la cellule à l'intersection de la ligne h et de la colonne f et son symétrique par rapport à la diagonale l'année t , ainsi que l'*odds ratio* généralisé anti-symétrique $GOR_{hf(t)}^A$ comme la moyenne géométrique du premier et de l'inverse du second. Alors, de manière analogue à la décomposition de l'association proposée plus haut, l'*odds ratio* généralisé peut être écrit comme le produit de ces deux quantités :

$$GOR_{hf(t)} = \sqrt{GOR_{hf(t)} GOR_{fh(t)}} \sqrt{\frac{GOR_{hf(t)}}{GOR_{fh(t)}}} = GOR_{hf(t)}^S GOR_{hf(t)}^A$$

En notant par souci de concision $\sigma_{hft}^{HFT} = s_{hf}^{HF} + \psi_t s_{hf}^{HFT}$ la spécification symétrique à forme de régression, l'*odds ratio* généralisé anti-symétrique s'écrit dans le dernier modèle, pour toute combinaison de h et de f telle que $p_h > p_f$ (configuration à hypergamie féminine) :

$$\begin{aligned}
 GOR_{hf(t)}^A &= \sqrt{\frac{GOR_{hf(t)}}{GOR_{fh(t)}}} \\
 &= \sqrt{\frac{\exp\left[\frac{HF}{(H-1)(F-1)}\left(\sigma_{hft}^{HFT} + \delta_t^{\mathcal{T}}\right)\right]}{\exp\left[\frac{HF}{(H-1)(F-1)}\sigma_{fht}^{HFT}\right]}} \\
 &= \exp\left[\frac{1}{2}\frac{HF}{(H-1)(F-1)}\left(\sigma_{hft}^{HFT} + \delta_t^{\mathcal{T}} - \sigma_{fht}^{HFT}\right)\right] \\
 &= \exp\left[\frac{1}{2}\frac{HF}{(H-1)(F-1)}\left(\sigma_{hft}^{HFT} + \delta_t^{\mathcal{T}} - \sigma_{hft}^{HFT}\right)\right] \\
 &= \exp\left[\frac{1}{2}\frac{HF}{(H-1)(F-1)}\delta_t^{\mathcal{T}}\right]
 \end{aligned}$$

On voit que, dans le modèle retenu, l'*odds ratio* généralisé anti-symétrique est déterminé uniquement par la valeur du paramètre associé aux configurations à hypergamie féminine $\delta_t^{\mathcal{T}}$ et par les dimensions de la table. Lorsque les configurations à hypogamie féminine sont prises comme ci-dessus pour référence, il mesure ainsi l'intensité générale de l'association une année donnée dans les configurations à hypergamie féminine par rapport à celle dans les configurations à hypogamie féminine. Sa formule est équivalente à celle de l'*odds ratio* généralisé, mais à partir du coefficient $\delta_t^{\mathcal{T}}$ divisé par 2 plutôt que du coefficient d'interaction classique. Pour faciliter l'interprétation, ce sont les valeurs de ce rapport que nous présentons, plutôt que le coefficient $\delta_t^{\mathcal{T}}$ lui-même. Le modèle avec évolution temporelle linéaire étant retenu, ce paramètre doit plus précisément être remplacé par $\delta + t\delta^{\mathcal{T}}$.

3.1.1 L'inversion de l'hypergamie relative de diplôme

Un premier modèle postule qu'il existe une tendance à l'hypergamie relative, mais que celle-ci est restée stable au cours du temps – l'association symétrique étant, comme dans tous les modèles, autorisée à varier suivant la spécification à forme de régression. Ce modèle améliore la description des données par rapport au modèle complètement symétrique d'après tous les indicateurs, mais de manière assez limitée : seuls 0,04 % des couples sont mieux classés par ce modèle. Un second modèle, autorisant l'hypergamie relative à varier de manière linéaire au cours du temps, apparaît meilleur selon le BIC et l'AIC, mais l'amélioration est si faible que l'indice de dissimilarité reste stable : cela n'est guère étonnant au vu de la faible importance de l'hypergamie relative pour la description des données. Finalement, un troisième modèle autorisant le paramètre d'hypergamie à varier de manière libre chaque année d'enquête ne se justifie pas au vu de sa complexité d'après le BIC et l'AIC.

C'est donc le modèle comprenant une tendance linéaire que nous retenons¹⁰. Celui-ci peut se résumer par deux paramètres, mesurant respectivement l'intensité de l'hypergamie relative au début et à la fin de la période (Figure 4.3). Ils décrivent une évolution très nette : en 1969, l'*odds ratio* généralisé anti-symétrique correspondant aux couples à hypergamie féminine était de 1,14, indiquant une association plus forte dans ces cellules que dans celles à hypogamie féminine ; en 2011, au contraire, il est de 0,97, indiquant une tendance légèrement inverse. L'hypergamie relative de diplôme, qui avait une certaine réalité au début de la période, s'est donc inversée au cours du temps, parallèlement à l'avènement de la supériorité éducative des femmes.

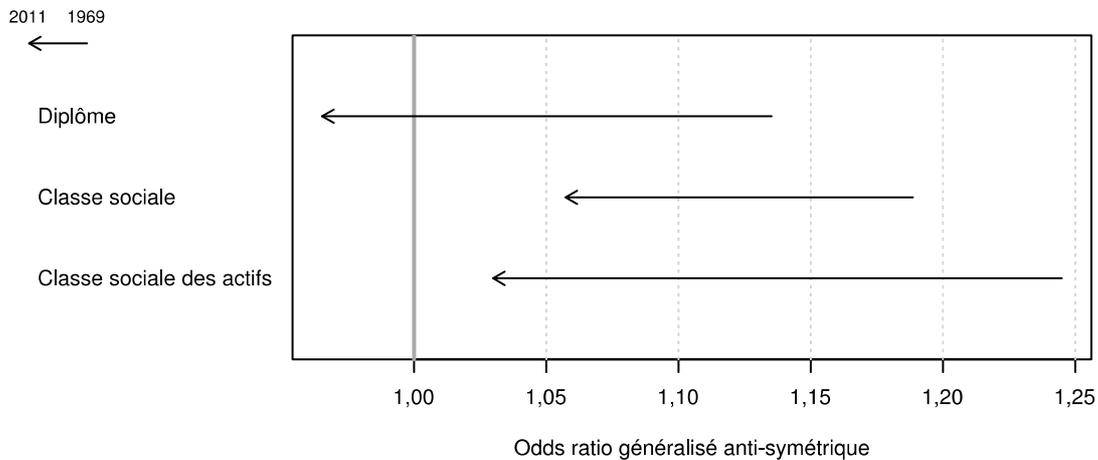
10 Nous avons vérifié au préalable que les séries de coefficients du modèle avec évolution temporelle libre ne présentaient pas de rupture de série trop marquée, qui fausserait la mesure linéaire.

Tableau 4.2 : Indicateurs d'ajustement des modèles

	D. L.	Déviante	Δ (%)	BIC	AIC
Diplôme :					
Indépendance	3 738	95 445	20,97	48 408	87 969
Quasi-symétrie stable	3 694	7 397	5,44	-39 087	9
Quasi-sym. à forme de régression	3 610	5 501	4,34	-39 925	-1 719
Stabilité de l'hypergamie	3 609	5 450	4,30	-39 964	-1 768
Évolution linéaire de l'hypergamie	3 608	5 359	4,30	-40 042	-1 857
Évolution libre de l'hypergamie	3 568	5 306	4,28	-39 592	-1 830
Classe sociale :					
Indépendance	3 738	81 124	18,69	33 940	73 648
Quasi-symétrie stable	3 694	9 268	5,26	-37 360	1 880
Quasi-sym. à forme de régression	3 610	7 682	4,46	-37 886	462
Stabilité de l'hypergamie	3 609	7 512	4,45	-38 044	294
Évolution linéaire de l'hypergamie	3 607	7 502	4,46	-38 028	288
Évolution linéaire de l'hypergamie, double	3 605	7 483	4,43	-38 022	273
Évolution libre de l'hypergamie	3 528	7 356	4,33	-37 178	300
Classe sociale des actifs :					
Indépendance	3 738	58 985	19,47	13 537	51 509
Quasi-symétrie stable	3 694	8 193	5,96	-36 720	805
Quasi-sym. à forme de régression	3 610	6 679	5,00	-37 213	-541
Stabilité de l'hypergamie	3 609	6 542	5,01	-37 337	-676
Évolution linéaire de l'hypergamie	3 607	6 516	5,01	-37 338	-698
Évolution linéaire de l'hypergamie, double	3 605	6 511	5,00	-37 320	-699
Évolution libre de l'hypergamie	3 528	6 405	4,87	-36 490	-651
Classe sociale d'origine :					
Indépendance	2 130	16 811	10,34	-9 348	12 551
Quasi-symétrie stable	2 095	2 720	3,88	-23 009	-1 470
Quasi-sym. à forme de régression	2 032	2 473	3,66	-22 482	-1 591
Stabilité de l'hypergamie	2 031	2 472	3,66	-22 470	-1 590
Évolution linéaire de l'hypergamie	2 029	2 472	3,65	-22 447	-1 586
Évolution libre de l'hypergamie	1 974	2 393	3,57	-21 850	-1 555

D. L. : degrés de liberté. Δ : indice de dissimilarité.

Figure 4.3 : Évolution de l'hypergamie relative de diplôme et de classe sociale entre 1969 et 2011



Source : enquêtes Emploi 1969-2011 (Insee)

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 30 à 59 ans (et les deux conjoints ont déjà travaillé dans le cas de la classe sociale)

L'*odds ratio* généralisé anti-symétrique (Encadré 4.1, p. 259) est égal au rapport entre les *odds ratios* généralisés correspondant aux configurations à hypergamie féminine et ceux correspondant aux configurations à hypogamie féminine. Une valeur supérieure à 1 indique une tendance à l'hypergamie féminine au-delà de ce qu'impose la structure de la population, alors qu'une valeur inférieure à 1 indique une tendance à l'hypogamie féminine.

Nous retrouvons dans le cas de la France l'observation plus générale d'A. Esteve, J. García-Román et I. Permanyer (2011) selon laquelle hypergamie absolue et hypergamie relative vont toujours dans le même sens. Il semble ainsi que les normes sociales, approchées ici par les choix que les individus font au-delà de ce que la disponibilité des conjoints exige, ne soient que le reflet des configurations de couple les plus fréquentes dans une cohorte donnée. En d'autres termes, le développement des couples dans lesquels la femme est plus diplômée que son conjoint – démographiquement inévitable du fait de l'élévation du niveau d'éducation des femmes – semble avoir entraîné une évolution des préférences des individus, ou des normes sociales intériorisées ou imposées à eux par le biais de leur entourage. La relative facilité avec laquelle les normes genrées paraissent ainsi avoir cédé et s'être retournées étonne : il semble que la domination masculine s'appuie avant tout sur des situations structurelles, les constructions normatives venant se surajouter à une situation de fait. L'hypothèse, faite par Hans-Peter Blossfeld et Andreas Timm (2003,

p. 10), selon laquelle l'hypergamie masculine aurait dû rester exceptionnelle du fait de la persistance de normes sociales à son encontre est ici clairement invalidée.

Le diplôme constitue-t-il une exception ? De fait, on a vu au chapitre 3 que l'hypergamie féminine en termes de salaires n'avait pas diminué. Comme on l'a déjà évoqué, la supériorité scolaire des filles peine à se traduire en termes professionnels. On pourrait donc penser que la norme d'hypergamie, ayant disparu du point de vue du diplôme, persiste cependant du point de vue de la classe sociale, qui, avec le salaire, constituerait le refuge des rôles de genre au sein du couple. Il n'en est rien.

3.1.2 La quasi-disparition de l'hypergamie relative de classe sociale

Nous avons déjà noté plus haut que l'hypergamie de classe sociale suivait globalement ce qui serait attendu en situation de mise en couple au hasard, impliquant une certaine faiblesse de la norme d'hypergamie. Cette interprétation est confirmée par les modèles (Tableau 4.2). Si le modèle de stabilité de l'hypergamie constitue une amélioration par rapport au modèle postulant une symétrie parfaite, la différence avec ce modèle est si faible que l'indice de dissimilarité varie à peine (ensemble des couples), voire augmente très légèrement (couples bi-actifs seulement)¹¹. Le modèle d'évolution linéaire de l'hypergamie améliore encore légèrement la description des données selon l'AIC, mais moins clairement selon le BIC. Un modèle comprenant une variation libre au cours du temps n'améliore pas la description des données, mais l'examen des coefficients annuels suggère cependant que des ruptures de séries importantes ont eu lieu en 1982 et en 2003, en particulier dans le premier cas lorsque les inactifs sont inclus dans le champ. Nous introduisons donc dans le modèle avec évolution linéaire des paramètres identifiant la série d'enquête, afin de ne pas tenir compte des sauts liés à ces ruptures¹². Ces paramètres améliorent le modèle selon l'AIC, mais pas selon le BIC.

C'est ce modèle plus correct que nous retenons (Figure 4.3). L'hypergamie relative de classe sociale était en 1969 légèrement supérieure à celle de diplôme, avec un *odds ratio* généralisé anti-symétrique de 1,19 à 1,24. Ce résultat est conforme à l'hypothèse avancée

11 Cette évolution paradoxale s'explique par le fait que l'estimation par le maximum de vraisemblance minimise la déviance, qui diffère légèrement de l'indice de dissimilarité. Cette discordance ne s'observe cependant que lorsque l'amélioration apportée par un paramètre est minime.

12 Le niveau de l'hypergamie est calculé en ajustant les séries d'enquêtes à la dernière (après 2003), que nous considérons comme parmi les plus fiables. La tendance temporelle est, elle, indépendante de ce choix.

dans la première section concernant la plus grande importance du statut professionnel dans la définition des rôles de genre. Mais cette hypergamie a très nettement diminué, au point de se limiter en 2011 à un *odds ratio* généralisé anti-symétrique de 1,03 à 1,06¹³.

Ce résultat est frappant : alors que dans le cas du diplôme, on pouvait penser que les comportements ne faisaient que prolonger la supériorité éducative des femmes, cette interprétation est exclue dans le cas de la classe sociale. En effet, comme on l'a observé plus haut (Figure 4.1, p. 256), les femmes sont toujours moins représentées que les hommes dans les emplois de cadres et professions intellectuelles, et les couples à hypergamie féminine continuent d'être nettement majoritaires. La quasi-disparition de l'hypergamie de classe sociale semble donc refléter une évolution du marché conjugal qui n'a pas son équivalent sur le marché du travail. En d'autres termes, la supériorité de l'homme en termes professionnels n'apparaît plus en elle-même comme un critère pour la formation des couples, ni pour les choix de carrière réalisés après la mise en couple ; en revanche, la ségrégation et la discrimination sur le marché du travail d'un côté, et les charges familiales que les femmes continuent d'assurer plus souvent que les hommes de l'autre, continuent d'entraver les carrières féminines. Ce paradoxe fait écho à la contradiction entre un idéal d'égalité des conjoints, qui a entraîné des effets bien réels, et la persistance d'inégalités entre hommes et femmes sur de nombreux plans, au premier rang desquels l'activité professionnelle et les tâches domestiques (Kaufmann, 1992).

3.1.3 L'introuvable hypergamie relative d'origine sociale

L'étude de l'hypergamie d'origine sociale confirme la faiblesse de cette norme. Ici, les paramètres d'hypergamie n'améliorent absolument pas le modèle postulant une association parfaitement symétrique d'après le BIC et l'AIC. En outre, les intervalles de confiance nous permettent d'exclure l'existence d'une hypergamie relative d'origine sociale d'un niveau autre que négligeable. Ce résultat, qui contraste avec la force de l'homogamie d'origine sociale mesurée au chapitre 2, achève de mettre en question la réalité d'une norme d'hypergamie féminine.

13 L'irrégularité des coefficients annuels du modèle avec évolution temporelle libre (non repris ici) nous incite à considérer avec prudence la différence apparente entre l'hypergamie observée parmi l'ensemble des couples et celle observée parmi les seuls couples bi-actifs.

Au total, les analyses présentées dans cette section dressent un tableau relativement complexe, mais convergent largement pour contredire l'importance de la norme d'hypergamie féminine. Inexistante du point de vue de l'origine sociale, l'hypergamie relative s'est inversée du point de vue du diplôme dès que la massification scolaire a mis fin à l'infériorité des diplômés féminins. Du point de vue de la classe sociale des conjoints, enfin, elle a presque complètement disparu en quarante ans, alors que l'égalité entre hommes et femmes sur le plan professionnel est loin d'être atteinte.

Ces résultats contrastent avec la résistance de l'hypergamie relative de salaire observée au chapitre 3. Cette différence pourrait s'interpréter comme la marque d'une plus grande proximité des dimensions considérées à ce qui constitue le cœur des rôles de genre au sein du couple : l'opposition entre le rôle de pourvoyeur de revenus et celui de responsable de la tenue du ménage. Suivant cette lecture, le diplôme ne constituerait qu'une dimension accessoire, de même que – dans une moindre mesure – la profession, puisque l'ajustement de l'investissement professionnel des femmes (notamment en termes de temps de travail) est toujours requis pour concilier activité professionnelle et tâches domestiques. Ainsi s'expliquerait la persistance de l'hypergamie relative de salaire parmi les couples bi-actifs. Pourtant, les évolutions du point de vue des professions exercées par les femmes et des salaires qu'elles en retirent, si elles sont moins spectaculaires que celles de la structure par diplôme, sont loin d'être négligeables. S'il est incontestable que l'hypergamie de salaire fait preuve d'une plus grande résistance que celle de diplôme, il n'est pas certain qu'il puisse en être ainsi indéfiniment.

4 L'évolution du célibat : une convergence entre hommes et femmes

Si le cadre d'interprétation décrit dans la première section de ce chapitre est correct, les évolutions majeures de l'hypergamie féminine que nous venons de mettre en évidence se sont nécessairement accompagnées d'une transformation profonde de la distribution sociale du célibat chez les deux sexes. On l'a vu, dans le modèle classique de l'hypergamie, les hommes aux positions les moins élevées et les femmes aux positions les plus élevées en termes de diplôme et de profession sont les groupes les plus frappés par le célibat, du fait d'un marché conjugal qui leur est défavorable. Étant donné que l'hypergamie féminine a

presque disparu selon la classe sociale, et s'est inversée selon le diplôme, on s'attendrait à ce que ces différences genrées se soient atténuées, se soient complètement résorbées, voire se soient inversées. L'ampleur des évolutions du célibat confirmera, ou nuancera, l'importance des modifications des rôles de genre au sein des couples.

Le célibat est ici défini, à partir des enquêtes Emploi, comme le négatif de la vie en couple : un individu est considéré comme célibataire s'il ne fait pas partie d'un couple. Rappelons que seuls les couples formés de la personne de référence du ménage et de son conjoint sont pris en compte, ce qui constitue une approximation acceptable de l'ensemble des couples dans les limites d'âges retenues (*cf.* chapitre 1, p. 77). Les unions non cohabitantes ne sont pas prises en compte, ce qui pourrait notamment tendre à augmenter le taux de célibat mesuré parmi les groupes les plus touchés par des difficultés économiques, cette question étant l'un des facteurs de la situation « ni seul ni en couple » (Régnier-Loilier, Beaujouan et Villeneuve-Gokalp, 2009, p. 103). Nous retiendrons au chapitre suivant une définition plus ouverte de la vie en couple qui lèvera cette limite.

Comme on l'a vu au chapitre 1, la série d'enquêtes 1969-74 fournit des estimations de la proportion de personnes en couple assez différentes des séries suivantes. Nous retenons donc l'année 1975 comme premier point.

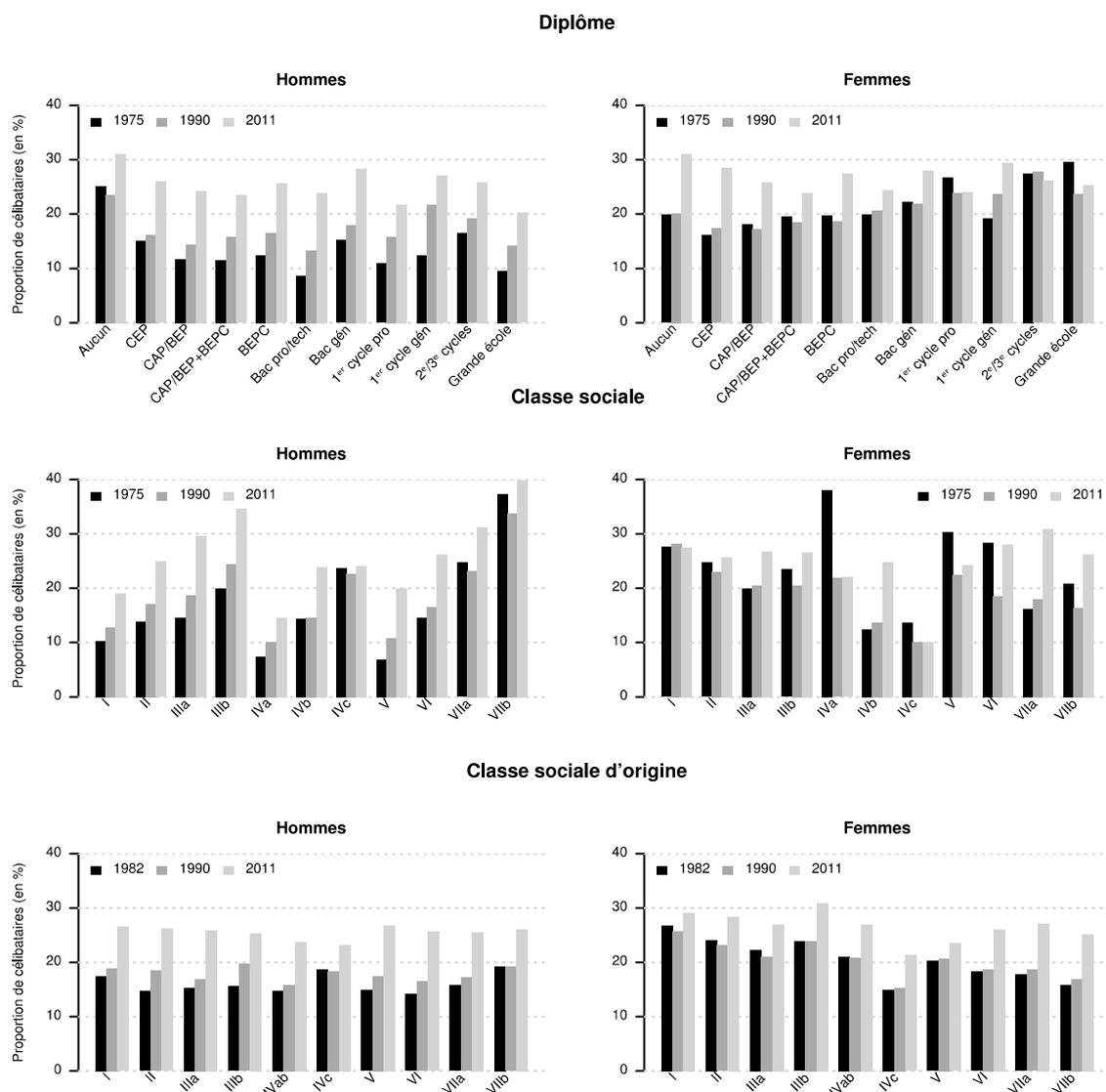
4.1 Une convergence entre hommes et femmes du célibat selon le diplôme

Comme on s'y attendait, les taux de célibat par diplôme se sont largement rapprochés entre hommes et femmes entre 1975 et 2011. Mais on est loin d'observer une inversion des différences genrées qui serait le pendant de celle de l'hypergamie féminine. La répartition du célibat en 1975 correspond bien au schéma classique déjà décrit (Figure 4.4) : chez les femmes, le célibat progressait de manière régulière avec le niveau de diplôme, alors que chez les hommes, seuls les moins diplômés se distinguaient par un célibat plus fréquent. Ce tableau était encore largement valable en 1990 (Albouy et Breuil-Genier, 2012) : la plus grande part des évolutions ont pris place au cours des vingt dernières années. Entre 1990 et 2011, en effet, le célibat féminin s'est développé principalement parmi les moins diplômées, au point que l'on n'observe plus guère de différences systématiques suivant le niveau de diplôme (Daguet et Niel, 2010 ; Prioux, Mazuy et Barbieri, 2011, p. 421). Les non diplômées

ont connu l'augmentation la plus forte du célibat, qui est resté au contraire très stable parmi les diplômées du supérieur¹⁴. Chez les hommes, le développement du célibat a été moins fort parmi les non diplômés, aboutissant à une légère réduction de l'écart avec les autres groupes. Au total, on relève en 2011 un schéma d'ensemble commun aux deux sexes : les non diplômés sont plus fréquemment célibataires que les autres groupes, avec un taux de célibat de 31 %. Quelques différences persistent, comme le faible célibat des hommes diplômés des grandes écoles (20 %, soit le plus faible de tous), qui contraste avec le taux plus proche de la moyenne des femmes de même diplôme (25 %).

14 Seules les diplômées du premier cycle général ont connu une hausse du célibat, qui correspond pour l'essentiel à un rattrapage des autres groupes de diplômées du supérieur.

Figure 4.4 : Proportion d'hommes et de femmes célibataires selon le diplôme, la classe et l'origine sociales en 1975 ou 1982, en 1990 et en 2011



Source : enquêtes Emploi 1975-2011 (Insee)

Champ : individus âgés de 30 à 59 ans (et ayant déjà travaillé dans le cas de la classe sociale)

Sont considérés comme célibataires les individus ne vivant pas en couple cohabitant à la date de l'enquête (couples comprenant la personne de référence du ménage uniquement).

Les proportions sont estimées à l'aide d'une régression locale de degré 1 (méthode de lissage non paramétrique, cf. légende de la figure 3.3, p. 202).

4.2 Des différences genrées moindres mais persistantes selon la classe sociale

Les évolutions sont moins marquées suivant la classe sociale. Chez les hommes, les mêmes traits se retrouvent au début et à la fin de la période, le développement du célibat ayant affecté de manière similaire la plupart des groupes. On notait ainsi une diminution du célibat à mesure que l'on progressait dans l'échelle de statut parmi les salariés, avec un taux de 37 % (40 % en 2011) parmi les ouvriers agricoles en 1975 et de seulement 10 % (19 % en 2011) parmi les cadres et professions intellectuelles supérieures (I). On peut noter que les techniciens et contremaîtres (V), et dans une moindre mesure les ouvriers qualifiés (VI), étaient moins fréquemment célibataires que leur position sur l'échelle sociale ne l'aurait supposé. Parmi les indépendants, les agriculteurs (IVc) se distinguaient par un taux de célibat élevé en 1975 (légèrement au-dessus de 20 %), mais qui, stable sur toute la période, apparaît faible relativement aux autres groupes en 2011¹⁵. Les indépendants avec salariés (IVa) présentent un taux de célibat très faible bien qu'en croissance sur toute la période, alors que ceux sans salariés (IVb) ont vu leur taux de célibat nettement augmenter depuis 1990.

Les différences entre groupes salariés sont moins marquées chez les femmes, mais les évolutions sont plus nettes. En 1975, l'échelle du célibat était approximativement l'inverse de celle relevée chez les hommes, quoique bien moins marquée : le célibat était plus fréquent à mesure que l'on s'élevait dans la hiérarchie, passant de 16 % pour les ouvrières non qualifiées (VIIa) à 28 % pour les cadres et professions intellectuelles supérieures (I). En miroir de ce que l'on notait chez les hommes, les techniciennes et contremaîtres (V), et les ouvrières qualifiées (VI) – groupes peu représentés chez les femmes –, connaissaient des taux de célibat élevés en 1975. De même, le célibat des agricultrices (IVc) était faible (autour de 10 %) – du fait en particulier de la faible proportion de femmes gérant seules une exploitation, liée elle-même à l'exode rural des filles d'agriculteurs (Bourdieu, 1962 ; Jegouzo, 1972 ; Courgeau et Lelièvre, 1986) –, alors que celui des indépendantes avec salariés (IVa) était élevé. Seules les employées non qualifiées (IIIb) échappaient en partie à cette inversion entre hommes et femmes, avec un taux de célibat relativement élevé (23 %).

15 Ce taux global masque des différences importantes entre petits et gros agriculteurs, seuls les premiers connaissant véritablement des difficultés de mise en couple (Jegouzo, 1972, p. 28, 1991, p. 52).

De manière analogue à l'évolution relevée suivant le diplôme, le célibat féminin s'est d'abord développé, entre 1990 et 2011, parmi les classes salariées qui vivaient alors le plus souvent en couple. Une certaine égalisation a donc pris place (Buisson et Daguet, 2012). On note même une certaine inversion de l'échelle : les femmes cadres et professions intellectuelles supérieures (I) vivent désormais un peu plus souvent en couple que les ouvrières qualifiées et non qualifiées (VI, VIIa), avec 27 % de célibataires, contre respectivement 28 % et 31 %. Pour autant, on n'observe pas – ou pas encore – l'échelle déjà commentée chez les hommes. La convergence en termes de célibat est donc ici moins nette que l'affaiblissement de l'hypergamie de classe sociale.

4.3 Une atténuation des différences selon la classe sociale d'origine chez les femmes

Les variations du taux de célibat selon la classe d'origine sont moins importantes que les précédentes, en particulier chez les hommes. Chez ces derniers, seuls les individus issus du monde agricole (agriculteurs, IVc ; ouvriers agricoles, VIIb), et ceux issus des classes supérieures (I), présentaient un taux de célibat plus élevé que les autres groupes. L'augmentation globale du taux de célibat a nettement modifié ce schéma, au point que les fils d'indépendants (IVabc) se distinguent désormais par un célibat moins fréquent que la moyenne, alors que les autres groupes présentent des taux similaires.

Les écarts sont plus marqués chez les femmes. Ils faisaient écho en 1982 à ceux déjà observés selon la classe sociale : le célibat augmentait avec le statut social, à l'exception des filles d'agriculteurs (IVc), moins souvent célibataires. Mises en relation avec la relative constance du célibat masculin à travers les groupes (agriculteurs mis à part), ces différences impliquent que c'étaient les filles qui vivaient le plus fréquemment en couple chez les travailleurs manuels, et notamment agricoles, alors que c'étaient les garçons qui étaient le moins souvent célibataires chez les professions en col blanc et les indépendants (hors agriculture).

Si le développement général du célibat entre 1990 et 2011 a eu tendance à atténuer ces écarts, ils persistent néanmoins en partie. Il semble que l'on observe indirectement, déformées par la mobilité sociale, les tendances déjà commentées plus haut. Le fait que l'origine sociale présente des effets plus nets sur les taux de célibat des femmes que sur ceux

des hommes semble indiquer que la mise en couple est plus affaire de normes sociales et moins de situation professionnelle chez les premières que chez les seconds. Selon cette interprétation, l'atténuation relative des écarts pourrait être ici encore l'indice d'une convergence des facteurs de mise en couple des hommes et des femmes.

Dans l'ensemble, l'hypothèse d'un lien entre évolution de l'hypergamie féminine et des taux de célibat se trouve confirmée. À l'inversion de l'hypergamie de diplôme correspond une convergence entre hommes et femmes des taux de célibat par niveau d'éducation. À la disparition de l'hypergamie de classe sociale fait écho l'atténuation des différences genrées suivant cette dimension. Enfin, l'absence d'hypergamie de classe sociale d'origine se traduit par des variations peu nettes du célibat par origine sociale chez les hommes, et des écarts moins nets que pour les autres dimensions chez les femmes. À chaque fois, le célibat suit les mêmes tendances temporelles que l'hypergamie, quoique de manière moins affirmée : on peut en conclure que l'absence de choix d'un conjoint, ou le choix de l'absence d'un conjoint, représente une prolongation du phénomène social du choix du conjoint.

À la différence de l'hypergamie, pour laquelle un sexe constitue par définition l'image inversée de l'autre, le célibat nous permet d'observer lequel des deux sexes a connu les évolutions les plus significatives. On assiste de ce point de vue à la convergence de la distribution du célibat des femmes par classe sociale, par classe sociale d'origine, et surtout par diplôme vers celle des hommes. En effet, l'échelle sociale du célibat est restée assez stable chez ces derniers, à l'exception de groupes bien particuliers comme les agriculteurs. L'augmentation du niveau d'éducation des femmes et de leur activité professionnelle a donc eu pour conséquence de rapprocher les facteurs de la vie en couple des femmes de ceux des hommes. Comme on l'a évoqué plus haut, cette évolution peut s'expliquer par une modification à la fois des comportements des hommes comme de ceux des femmes.

5 Une synthèse : le modèle d'association anti-symétrique

L'étude séparée des évolutions de l'hypergamie et du célibat a confirmé la relation qu'entretiennent ces deux phénomènes, visible aux tendances parallèles qu'ils suivent. Pour conclure cette analyse, nous chercherons dans cette section à synthétiser ces deux faces

d'une même pièce – la détermination genrée du choix ou du non-choix du conjoint – dans une représentation commune. Alors que nous avons mesuré plus haut l'hypergamie en la définissant à partir d'un classement extérieur au choix du conjoint établi *a priori*, il s'agira ici de décrire les asymétries entre hommes et femmes telles qu'elles apparaissent dans l'association mesurée à partir des tables d'homogamie. Cette représentation libre pourra être confrontée aux résultats déjà obtenus, et fournir une épreuve supplémentaire au cadre théorique que nous avons élaboré.

Les tables d'homogamie utilisées dans cette section combinent celles retenues jusqu'ici avec la distribution du célibat entre les groupes sociaux. Pour cela, l'absence de conjoint est considérée comme un choix supplémentaire possible, ou encore comme un type de conjoint¹⁶. Les pourcentages en lignes et en colonnes de cette table augmentée décrivent comme on s'y attendrait la répartition des individus de chaque sexe selon les différentes situations possibles. Seule particularité, la cellule correspondant à l'absence de conjoint chez les deux sexes, dépourvue de signification, est nécessairement exclue de l'analyse – comme d'ailleurs l'ensemble des autres cellules de la diagonale (*cf. supra*).

Analyser l'hypergamie à partir d'une telle table suivant la démarche retenue plus haut ne semble pas l'approche la plus féconde : en effet, elle exigerait de considérer *a priori* le célibat comme étant le plus fréquent pour les individus d'un statut social élevé pour un sexe, et inversement pour l'autre. Ce postulat reviendrait à supposer, de manière quelque peu tautologique, l'hypothèse dont on cherche justement à évaluer la validité.

5.1 Le modèle d'association anti-symétrique

Un modèle peu utilisé est parfaitement adapté à notre objectif. Le modèle d'association lignes-colonnes anti-symétrique élaboré par Peter van der Heijden et Ab Mooijaart (1995) reprend le principe d'une description sans aucun *a priori* de l'association suivi par les modèles RC utilisés au chapitre 2, mais pour décrire la partie anti-symétrique de l'association (*cf.* Encadré 4.1, p. 259), autrement dit l'hypergamie. Ce modèle cherche à résumer de la meilleure manière possible les écarts à l'association symétrique par une ou plusieurs paires de dimensions – qui ne font pas sens prises séparément –, sur lesquelles les catégories en lignes et en colonnes partagent les mêmes coordonnées.

16 Nous tenons à remercier Thomas Amossé pour nous avoir souligné l'intérêt d'une telle approche.

Suivant la décomposition de l'association en une partie symétrique et une partie anti-symétrique présentée plus haut, il est possible de mesurer l'importance de l'asymétrie hommes-femmes décrite chaque année par une paire de dimensions au coefficient d'association intrinsèque anti-symétrique et au MAOR anti-symétrique qui lui sont associés. Ces indices sont définis de manière similaire aux versions introduites au chapitre 2, mais en tenant compte uniquement de la composante anti-symétrique de l'association. Le MAOR anti-symétrique est ainsi défini comme l'écart-type de la moyenne (géométrique) de tous les *odds ratios* de la table multipliés par l'inverse de leur symétrique. Il s'interprète donc comme une mesure de l'ampleur des écarts à la symétrie parfaite des *odds ratios*. Comme dans les modèles RC-L, il peut éventuellement être inférieur à 1 : il signale alors l'inversion d'une paire de dimensions, les configurations jusqu'ici sous-représentées par rapport à leur symétrique devenant alors les plus fréquentes. Le choix de fixer le MAOR à une valeur supérieure à 1 une année donnée est purement conventionnel : sans rien modifier à l'association décrite, il est toujours possible de remplacer le MAOR par son inverse en intervertissant les deux axes d'une paire. Le détail des démonstrations relatives à ces variantes des deux indicateurs est développé en annexe B (p. 525).

Contrairement aux modèles RC et à l'analyse des correspondances, les plans représentant les positions des catégories sur une paire de dimensions ne se lisent pas aux oppositions sur chaque axe (coordonnées cartésiennes), mais de manière orientée (coordonnées polaires), en suivant ou en allant à l'inverse du sens des aiguilles d'une montre. C'est le sens de l'angle formé par un point, l'origine des axes, et un autre point, qui indique si les configurations dans lesquelles l'homme appartient à la première catégorie et la femme à la seconde sont sur- ou sous-représentées (en termes d'*odds ratio* généralisé anti-symétrique) par rapport à leur symétrique. En prenant pour référence une catégorie donnée, les catégories d'appartenance des conjoints se trouvent divisées en deux groupes : dans le sens horaire, et jusqu'à 180°, se trouvent les groupes plus fréquemment associés aux hommes de la catégorie ; dans le sens inverse, ceux plus fréquemment associés aux femmes de la catégorie. Enfin, un groupe connaît son asymétrie maximale avec les groupes dont les points forment un angle droit avec le sien ; il présente au contraire une association symétrique et des choix similaires avec les points qui forment un angle nul avec le sien ; et enfin une association symétrique, mais des choix opposés, avec les points qui forment un angle de 180° avec le sien (ceux qui lui sont diamétralement opposés).

Ces critères sont résumés par une figure plus intuitive à manier : plus l'aire du triangle reliant deux points à l'origine est grande, plus deux points présentent une forte asymétrie hommes-femmes ; la comparaison du sens de l'angle avec celui indiqué par la flèche donne le sens de l'asymétrie. Ainsi, les points les plus éloignés de l'origine des axes correspondent aux catégories présentant les asymétries les plus fortes entre hommes et femmes, alors qu'un point situé sur l'origine ne présente aucune asymétrie. Mais l'angle du segment reliant les points à l'origine n'a aucune signification en lui-même : il n'a de sens que relativement aux autres points. Le plan peut ainsi subir des rotations arbitraires sans modifier l'association décrite.

Le modèle est estimé à l'aide de notre paquet R `logmult` (Bouchet-Valat, 2015a). Étant donné l'ampleur limitée des asymétries à décrire (*cf. infra*) et la relative complexité du modèle, nous ne retenons qu'une seule paire de dimensions. À notre connaissance, le modèle comprenant plusieurs paires de dimensions n'a jamais été utilisé.

Finalement, notons que nous ne retenons pas ici la classe sociale des seuls actifs, puisque cette restriction ne fait pas sens pour les célibataires, et entraînerait une sur-représentation importante de ces individus s'ils étaient tous retenus dans le champ.

Encadré 4.2 : Équation du modèle d'association anti-symétrique

Le modèle d'association lignes-colonnes anti-symétrique utilisé ici est une extension à des tables à trois dimensions du modèle proposé par P. van der Heijden et A. Mooijaart (1995 ; voir aussi de Falguerolles et van der Heijden, 2002). Fondée sur une décomposition de Gower (Constantine et Gower, 1978, p. 301-302 ; Greenacre, 2000), cette spécification de l'association est similaire à celle du modèle d'association lignes-colonnes RC, mais vise à décrire la composante anti-symétrique de l'association définie à l'encadré 4.1 (p. 259).

Nous ne retenons ici qu'une seule paire de dimensions anti-symétriques. Comme précédemment, le modèle de base est le modèle quasi-symétrique à forme de régression. L'équation complète du modèle s'écrit :

$$\log m_{hft} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_t^T + \lambda_{ht}^{HT} + \lambda_{ft}^{FT} + s_{hf}^{HF} + \psi_t s_{hf}^{HFT} + \phi_t^A (v_h \mu_f - \mu_h v_f)$$

Comme dans le modèle RC, les paramètres μ et ν représentent les scores, et ϕ^A l'intensité de l'association. Mais ici, les scores des lignes et des colonnes sont identiques : les μ représentent les coordonnées des catégories sur la première dimension, et les ν celles sur la seconde dimension de la paire. Le coefficient d'association intrinsèque anti-symétrique ϕ^A s'applique aux deux dimensions. Pour être identifiables, les scores sont centrés et réduits en retenant une pondération par la moyenne des marges de la table :

$$\sum_{k=1}^K \mu_k \frac{P_{+k} + P_{k+}}{2} = \sum_{k=1}^K \nu_k \frac{P_{+k} + P_{k+}}{2} = 0$$

$$\sum_{k=1}^K \mu_k^2 \frac{P_{+k} + P_{k+}}{2} = \sum_{k=1}^K \nu_k^2 \frac{P_{+k} + P_{k+}}{2} = 1 \quad \text{avec } K = I = J$$

Sous ce modèle, l'*odds ratio* généralisé anti-symétrique varie selon la cellule considérée : il dépend des dimensions de la table et de la spécification anti-symétrique dont l'interprétation est développée dans le corps du texte. Soit, en notant $\sigma_{hft}^{HFT} = s_{hf}^{HF} + \psi_t s_{hf}^{HFT}$:

$$\begin{aligned} GOR_{hf(t)}^A &= \sqrt{\frac{GOR_{hf(t)}}{GOR_{fh(t)}}} \\ &= \sqrt{\frac{\exp\left[\frac{HF}{(H-1)(F-1)} \left(\sigma_{hft}^{HFT} + \phi_t^A (v_h \mu_f - \mu_h \nu_f)\right)\right]}{\exp\left[\frac{HF}{(H-1)(F-1)} \left(\sigma_{fht}^{HFT} + \phi_t^A (v_f \mu_h - \mu_f \nu_h)\right)\right]}} \\ &= \exp\left[\frac{HF}{(H-1)(F-1)} \phi_t^A (v_h \mu_f - \mu_h \nu_f)\right] \end{aligned}$$

5.2 L'inversion des asymétries selon le diplôme

Comme les modèles décrivant l'hypergamie utilisés plus haut, les modèles d'association anti-symétrique améliorent clairement les modèles postulant une symétrie parfaite de l'association, mais de manière relativement modeste. Ce résultat confirme encore une fois l'importance limitée des asymétries hommes-femmes du point de vue du choix du conjoint, même en tenant compte de la possibilité du célibat.

L'ajout d'une paire de dimensions anti-symétrique au modèle de quasi-symétrie à forme de régression permet de classer correctement un peu moins de 0,5 % des effectifs supplémentaires (Tableau 4.3). Les indicateurs BIC et AIC indiquent nettement que cette amélioration est significative. L'évolution temporelle de cette association a elle aussi été notable, puisqu'elle permet de réduire l'indice de dissimilarité d'encore 0,6 points de pourcentage. Ici encore, le BIC et l'AIC préfèrent ce modèle plus riche.

La paire de dimensions identifiée (Figure 4.5) dessine un plan clairement divisé en deux parties. Sa moitié droite contient exclusivement la catégorie des célibataires (« Aucun conjoint ») ; sa moitié gauche déroule l'échelle des diplômes dans le sens inverse des aiguilles d'une montre. Ainsi, parmi les individus non diplômés, les hommes présentent une tendance au célibat bien plus forte que les femmes, l'angle droit entre les deux catégories correspondant à l'asymétrie la plus forte possible dans le sens horaire. En revanche, les hommes non diplômés se mettent plus rarement en couple que les femmes du même groupe avec un conjoint diplômé (sens anti-horaire). En s'élevant dans l'échelle des diplômes, on retrouve une opposition similaire : les femmes tendent à prendre des conjoints plus diplômés, les hommes des conjointes moins diplômées. Mais au-delà du CAP et du BEP, ce sont les femmes qui ont le plus de chances d'être célibataires, l'angle avec la catégorie « Aucun conjoint » passant du sens horaire au sens anti-horaire. La différence entre célibat féminine et célibat masculin atteint son maximum parmi les diplômés du supérieur, dont les points forment approximativement un angle droit avec la catégorie « Aucun conjoint ». À considérer l'ensemble des asymétries entre hommes et femmes (célibat et hypergamie), on remarque que les catégories les plus éloignées de l'origine sont celles relevant du supérieur long (2^e/3^e cycles, grandes écoles) et le célibat. À l'inverse, les titulaires du certificat d'études, du BEPC ou d'un baccalauréat général présentent assez peu de différences entre hommes et femmes.

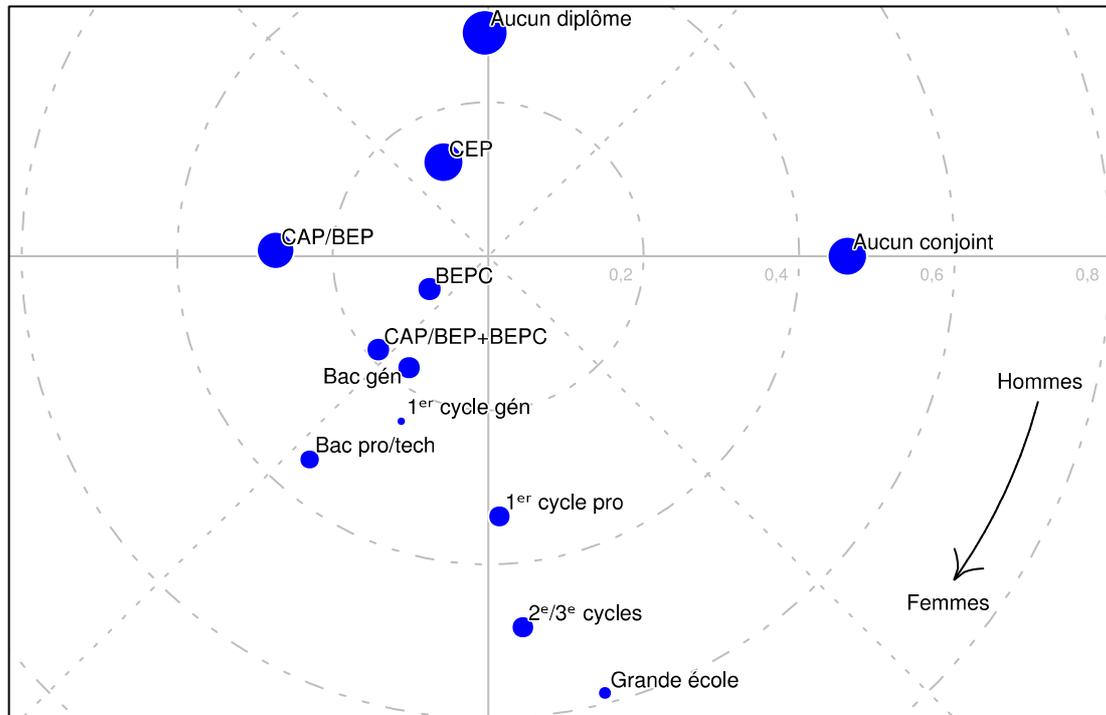
Tableau 4.3 : Indicateurs d'ajustement des modèles incluant les célibataires

	D. L.	Déviante	Δ (%)	BIC	AIC
Diplôme :					
Indépendance	5 029	217 813	20,31	150 649	207 755
Quasi-symétrie	4 964	17 958	5,65	-48 337	8 030
Quasi-sym. à forme de régression	4 859	11 075	4,46	-53 818	1 357
Stabilité de l'association anti-symétrique	4 840	9 133	4,02	-55 506	-547
Évolution de l'association anti-symétrique	4 799	7 106	3,43	-56 986	-2 492
Classe sociale :					
Indépendance	5 029	244 277	21,82	177 752	234 219
Quasi-symétrie	4 964	21 655	6,17	-44 010	11 727
Quasi-sym. à forme de régression	4 859	17 176	5,28	-47 101	7 458
Stabilité de l'association anti-symétrique	4 840	9 087	3,80	-54 939	-593
Évolution de l'association anti-symétrique	4 799	8 318	3,62	-55 165	-1 280
Classe sociale d'origine :					
Indépendance	2 960	65 276	12,70	26 872	59 356
Quasi-symétrie	2 906	4 852	3,80	-32 852	-960
Quasi-sym. à forme de régression	2 824	3 688	3,25	-32 951	-1 960
Stabilité de l'association anti-symétrique	2 807	3 298	3,02	-33 121	-2 316
Évolution de l'association anti-symétrique	2 778	3 198	2,95	-32 845	-2 358

D. L. : degrés de liberté. Δ : indice de dissimilarité.

Ces résultats illustrent bien le lien entre hypergamie féminine et distribution genrée du célibat. Ils mettent clairement en évidence que le célibat a des caractéristiques opposées chez les hommes et les femmes : chez les premiers, il est le fait de ceux qui ne peuvent trouver une conjointe moins diplômée qu'eux ; chez les secondes, de celles qui ne peuvent trouver de conjoint plus diplômé qu'elles. Rappelons que le modèle d'association anti-symétrique construit ce plan sans aucune connaissance de la hiérarchie des diplômes, ni de la nature particulière du célibat, dont la spécificité apparaît à partir des seules données. L'asymétrie hommes-femmes ainsi décrite correspond pourtant de très près au modèle classique de l'hypergamie féminine que nous avons déjà étudié plus haut avec une définition *a priori* des configurations à hypergamie et à hypogamie féminine.

Figure 4.5 : La paire de dimensions anti-symétrique pour le diplôme

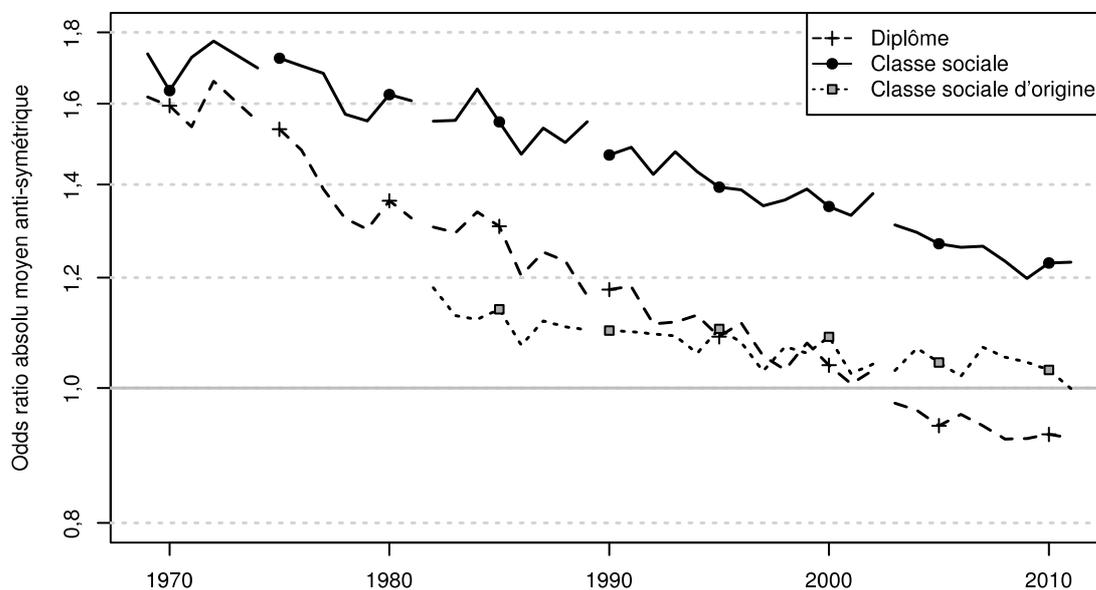


Source : enquêtes Emploi 1969-2011 (Insee)

Champ : individus âgés de 30 à 59 ans

Une pondération par la moyenne des marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes.

En cohérence avec les résultats déjà obtenus plus haut, ce schéma n'est pas resté stable au cours des quarante dernières années (Figure 4.6). Si la description du plan que nous venons de détailler s'appliquait bien au début de la période, l'intensité de la paire de dimension a continûment diminué depuis 1969, au point de s'inverser après 2000 : l'*odds ratio* généralisé anti-symétrique passe ainsi de 1,6 à 0,9. Ce résultat confirme donc l'inversion de l'hypergamie féminine déjà observée plus haut, tout en l'élargissant par la prise en compte simultanée du célibat.

Figure 4.6 : Évolution de l'intensité des paires de dimensions anti-symétriques

Source : enquêtes Emploi 1969-2011 (Insee)

Champ : individus âgés de 30 à 59 ans (et ayant déjà travaillé dans le cas de la classe sociale)

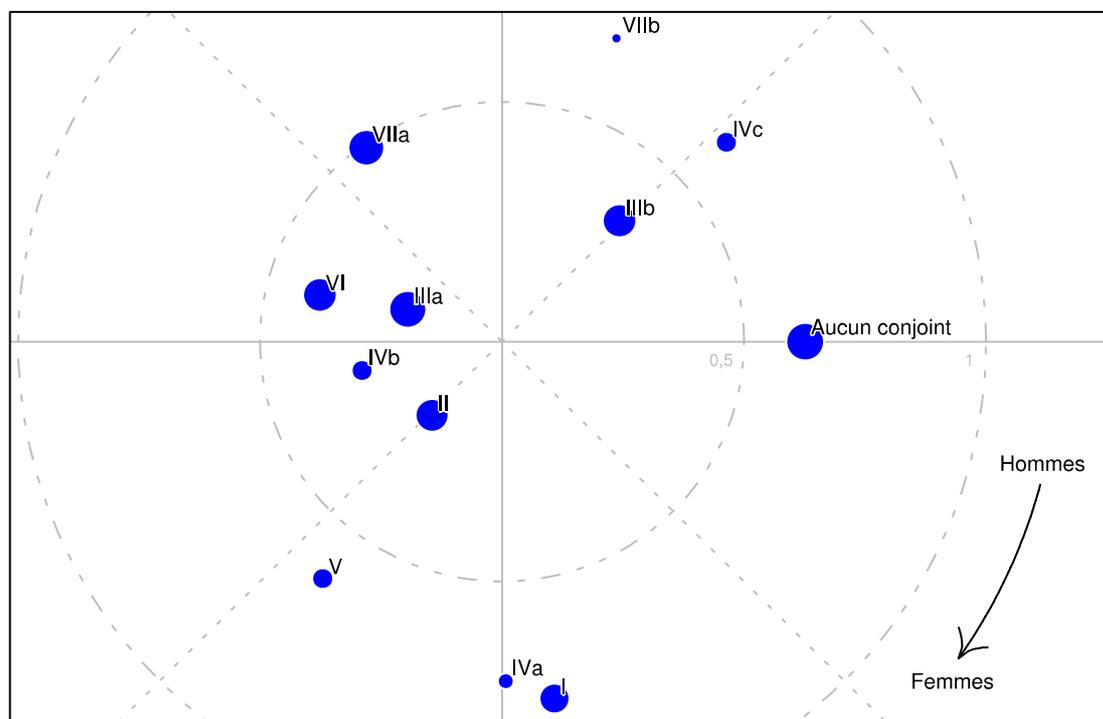
Un *odds ratio* absolu moyen anti-symétrique (cf. p. 276) de 1 correspond à une absence d'asymétrie hommes-femmes (au-delà des contraintes structurelles) selon la paire de dimensions considérée ; une valeur inférieure à 1 indique l'inversion du sens de l'asymétrie par rapport à ce qui est décrit sur le plan (Figure 4.5 pour le diplôme). L'échelle verticale logarithmique permet la comparaison des écarts entre valeurs (équivalents à des rapports entre *odds ratios*), en particulier entre valeurs supérieures et inférieures à 1. Une pondération par la moyenne des marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes.

5.3 La disparition des asymétries selon la classe sociale

L'application du modèle d'association anti-symétrique à la classe sociale confirme elle aussi les résultats déjà obtenus. Ce modèle permet de classer correctement près de 1,5 % des couples en plus par rapport à celui supposant une symétrie parfaite de l'association, soit nettement plus que pour le diplôme (Tableau 4.3, p. 280). En revanche, les évolutions temporelles apparaissent plus limitées, puisqu'elles n'améliorent ce classement que de 0,2 points de pourcentage. Le BIC et l'AIC confirment l'existence d'asymétries, ainsi que de leur évolution temporelle.

Le plan du modèle incluant une telle évolution se divise de manière relativement différente de celui concernant le diplôme : alors que la moitié gauche suit approximativement une échelle sociale, la moitié droite suit une logique bien plus complexe. Comme pour le diplôme, on observe qu'au sein des groupes au statut le plus élevé (cadres et professions intellectuelles supérieures, I ; indépendants avec salariés, IVa), les femmes sont plus souvent célibataires, et les hommes plus souvent hypogames. Mais l'échelle sociale qui se dessine en suivant le sens des aiguilles d'une montre est atypique. Ainsi, les techniciens et contremaîtres (V) apparaissent avant les cadres et professions intellectuelles de rang inférieur (II). Mais ce sont les groupes qui apparaissent en dernier, juste avant la catégorie des célibataires, qui brisent la hiérarchie : professions agricoles (agriculteurs, IVc ; ouvriers agricoles, VIIb) et employés non qualifiés (IIIb) se trouvent bien après les ouvriers non qualifiés (VIIa) et les employés qualifiés (IIIa). Cette position reflète un célibat plus fréquent parmi les hommes que parmi les femmes de ces catégories (notamment dans le monde agricole), on voit maintenant qu'il va de pair avec une tendance à l'hyogamie féminine¹⁷. En effet, les configurations dans lesquelles l'homme appartient à ces groupes, et sa conjointe aux cadres et professions intellectuelles (I, II) sont sur-représentées en termes d'*odds ratios* par rapport aux configurations inverses. Au contraire, les configurations dans lesquelles les conjointes sont ouvrières (VI, VIIa) sont sous-représentées. Cette particularité se retrouve sur le plan à l'opposition entre moitié supérieure gauche et moitié inférieure droite.

17 Cette interprétation est confirmée par le fait que la position de ces groupes au bas de la hiérarchie persiste lorsque le modèle est estimé sur la table excluant les célibataires (voir Bouchet-Valat, 2015b pour une analyse fondée sur la PCS à deux chiffres).

Figure 4.7 : La paire de dimensions anti-symétrique pour la classe sociale

Source : enquêtes Emploi 1969-2011 (Insee)

Champ : individus âgés de 30 à 59 ans et ayant déjà travaillé

Une pondération par la moyenne des marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes.

La différence de position entre ouvriers non qualifiés (VIIa) et employés non qualifiés (IIIb) est frappante. Alors que les ouvriers non qualifiés (VIIa) évitent de prendre pour conjointes des femmes placées plus haut qu'eux dans la hiérarchie des professions, et se mettent très fréquemment en couple avec des employées non qualifiées (IIIb), les hommes de ce dernier groupe ne font que rarement le choix inverse de se mettre en couple avec des ouvrières non qualifiées. L'asymétrie entre ces deux configurations va bien au-delà de ce que la structure de la population exige. Puisqu'elle ne peut s'expliquer par les situations respectives des ouvriers et employés non qualifiés sur le marché du travail, relativement similaires (Amossé et Chardon, 2006), elle semble devoir être interprétée en termes d'interaction entre identités professionnelles et rôles de genre. En effet, les employés non qualifiés sont parmi les professions les plus féminisées, qui ne se prêtent sans doute pas au maintien d'une identité masculine très affirmée, qui passerait notamment par le respect de la

norme d'hypergamie féminine. Ils travaillent aussi plus souvent dans de petites structures dans lesquelles la chaîne hiérarchique est plus courte (voire absente) et les normes de travail peu formalisées (Gollac, 1989 ; Burnod et Chenu, 2001), ce qui pourrait faciliter les rencontres allant à l'encontre de l'hypergamie féminine en rendant ce critère moins visible.

Ces hypothèses sont confortées par les résultats obtenus à partir de la catégorie socio-professionnelle à deux chiffres, qui montrent que l'hypogamie féminine est le fait des catégories les moins inscrites dans l'organisation hiérarchique du travail, comme les professions libérales, les professions de l'information, des arts et des spectacles, les professeurs, les instituteurs, les professions intermédiaires de la santé et du travail social, ou encore les ouvriers non qualifiés de type artisanal et les employés de commerce (Bouchet-Valat, 2015b). En outre, les hommes exerçant dans les professions très féminisées (en particulier les employés et les enseignants) déclarent plus souvent que les autres avoir rencontré leur conjointe sur leur lieu de travail (Bozon et Héran, 1988, p. 124), ce qui laisse potentiellement à l'organisation du travail une plus grande place pour jouer sur la mise en couple.

Ces pistes d'interprétation demandent à être confirmées par des analyses plus fines concernant notamment le cadre de travail de ces groupes ainsi que les lieux et conditions de la rencontre des conjoints. Elles pointent cependant l'importance de la hiérarchie – et, plus encore, de l'organisation du travail – dans la genèse de l'hypergamie féminine, à l'encontre de l'idéal romantique de la rencontre abstraite de toute influence sociale et d'égalité entre conjoints.

Pour autant, ce tableau n'est pas non plus resté intangible au cours des quarante dernières années (Figure 4.6). Les asymétries hommes-femmes en termes de classe sociale ont clairement décliné, l'*odds ratio* généralisé anti-symétrique passant de près de 1,8 à 1,2. Ce résultat est cohérent avec la très nette diminution de l'hypergamie féminine observée plus haut et l'atténuation un peu moins marquée des différences genrées de célibat par classe sociale. La tendance ne semble pas s'essouffler, indiquant une possible disparition complète des asymétries, voire leur inversion, à relativement brève échéance.

5.4 L'affaiblissement des asymétries selon la classe sociale d'origine

Finalement, le modèle d'association anti-symétrique permet de mettre en évidence des asymétries hommes-femmes qui étaient restées invisibles plus haut lors de l'utilisation d'une hiérarchie *a priori* des classes sociales. L'un des intérêts de ce modèle est en effet de mettre au jour les asymétries quelle que soit leur structure. (Tableau 4.3). Cependant, le modèle n'améliore pas aussi clairement la description des données que dans le cas du diplôme et de la classe sociale des conjoints. L'ajout d'une spécification anti-symétrique stable permet de classer correctement un supplément de 0,2 % des cas, et l'évolution temporelle, 0,05 % seulement. La significativité de ces améliorations est cependant attestée par la diminution du BIC (dans le premier cas uniquement) et de l'AIC (dans les deux cas).

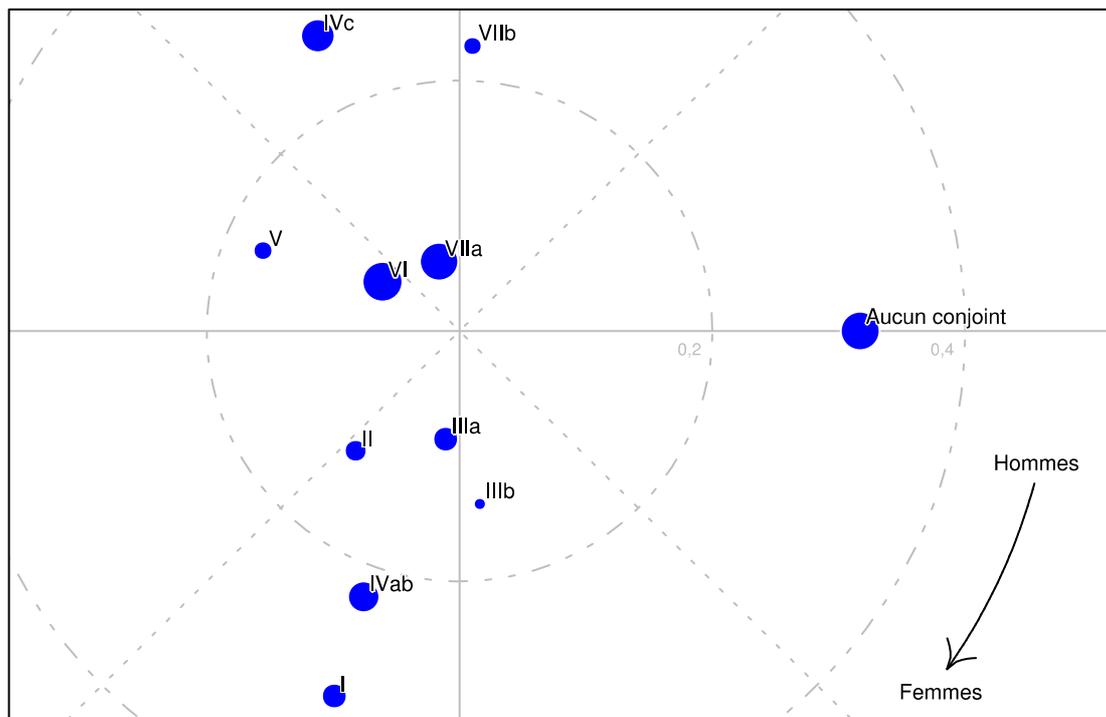
Le plan décrit par la paire de dimensions anti-symétrique est relativement similaire à celui déjà obtenu pour la classe sociale des conjoints (Figure 4.8). On retrouve, dans sa moitié gauche, l'échelle de statut déroulée approximativement dans le sens des aiguilles d'une montre. Néanmoins, les enfants d'employés (IIIa, IIIb) viennent bien avant ceux d'ouvriers (VI, VIIab) et de techniciens et contremaîtres (V) – ce qui contraste nettement avec la position des employés non qualifiés (IIIb), en bas de la hiérarchie, obtenue du point de vue de la classe sociale des conjoints. Cette différence entre effet de l'origine sociale et effet de la profession occupée peut s'expliquer par le fait que l'on compare ici les professions de deux hommes (les pères), alors que l'analyse à partir de la classe sociale des intéressés met en regard celle d'un homme et d'une femme. Du fait de la ségrégation genrée des professions, les hommes issus de ce groupe sont rarement devenus eux-mêmes employés non qualifiés, contrairement aux femmes de même condition (Vallet, 1999, p. 27, 29).

Mais cette explication n'est que partielle. En effet, cette hiérarchie diffère aussi de celle mise en évidence pour l'homogamie suivant la classe d'origine au chapitre 2 à l'aide du modèle RC-L (Figure 2.10, p. 177). Pris globalement, le plan oppose avant tout trois ensembles : les cols blancs (ainsi que les indépendants), les professions manuelles, puis les professions agricoles. Il recoupe ainsi la distinction selon le taux de célibat déjà identifiée plus haut. L'inversion des positions respectives des enfants d'employés non qualifiés (IIIb) et de techniciens et contremaîtres (V) illustre le plus nettement cette opposition entre cols blancs et cols bleus. Il semble bien que l'on soit ici face à des univers sociaux aux normes de genre différentes, qui ne sont que partiellement corrélées à leur statut social. Une fois de

plus, on remarque donc que les mécanismes expliquant les asymétries de genre dans la mise en couple diffèrent de ceux déterminant la position des groupes sur l'échelle sociale.

L'évolution temporelle des asymétries hommes-femmes selon la classe sociale d'origine est moins marquée que selon les deux autres dimensions (Figure 4.6, p. 282), ce qui est cohérent avec la faible amélioration apportée par le modèle. Néanmoins, elle indique une baisse relativement régulière de l'importance de la paire de dimensions, qui disparaît même en 2011. Moins clairement significative, cette tendance est donc néanmoins parfaitement cohérente avec les deux précédentes.

Figure 4.8 : La paire de dimensions anti-symétrique pour la classe sociale d'origine



Source : enquêtes Emploi 1982-2011 (Insee)

Champ : individus âgés de 30 à 59 ans

Une pondération par la moyenne des marges est utilisée, et la taille des points est proportionnelle aux effectifs moyens des catégories correspondantes.

6 Conclusion

Les évolutions de l'hypergamie de diplôme, de classe et d'origine sociales ont été majeures en France depuis 1969. La supériorité scolaire des filles s'est récemment traduite par une supériorité des femmes par le diplôme au sein du couple, non seulement en termes d'une plus grande fréquence des couples à hypogamie féminine, mais aussi au-delà des contraintes imposées par la structure par diplôme de la population (hypergamie relative) : désormais, comme dans nombre de pays du monde, le choix du conjoint tend à renforcer la tendance structurelle à former des couples dans lesquels la femme est la plus diplômée. Si l'évolution a été moins spectaculaire du point de vue de la classe sociale des conjoints, elle est cependant notable. Les femmes occupent toujours moins fréquemment que les hommes des positions professionnelles élevées, mais l'écart s'est atténué au cours du temps. Parallèlement à ce mouvement, l'hypergamie relative de classe sociale s'est fortement réduite, et pourrait à ce rythme disparaître dans les années à venir. Enfin, les asymétries genrées de choix du conjoint selon la classe sociale d'origine, qui ne trouvent par définition aucune justification dans la structure de la population, tendent elles aussi à disparaître.

Comme nous en avons fait l'hypothèse, ces évolutions ont eu pour conséquence une diminution des différences genrées dans la distribution sociale du célibat. L'inversion de l'hypergamie de diplôme s'est ainsi traduite par une convergence entre hommes et femmes de l'association entre taux de célibat et niveau d'éducation : les moins diplômés sont désormais le groupe le plus affecté par le célibat chez les deux sexes, alors que ce sont les plus diplômées qui étaient dans cette situation chez les femmes encore en 1990. Cette convergence apparaît moins nette du point de vue de la classe sociale, mais l'évolution est bien réelle : en 2011, les femmes cadres et professions intellectuelles supérieures vivent un peu plus souvent en couple que les ouvrières non qualifiées, se rapprochant en partie du schéma valable de longue date chez les hommes. Finalement, ce rapprochement partiel se retrouve dans la distribution du célibat par classe sociale d'origine : ici, les écarts entre femmes de différentes origines se sont réduits, donnant une image plus proche de celle observée parmi les hommes.

Enfin, le modèle d'association anti-symétrique nous a permis d'obtenir une vision d'ensemble des asymétries entre hommes et femmes du point de vue du choix du conjoint aussi bien que du célibat. Cette description a confirmé la structure commune de

l'hypergamie féminine et des différences genrées dans le célibat : dans l'ensemble, l'absence de conjointe est pour les hommes l'équivalent d'une conjointe située tout en bas de l'échelle sociale, alors qu'elle correspond pour les femmes à un conjoint situé à son sommet. Le mécanisme à l'origine du contraste entre célibat des hommes et célibat des femmes apparaît nettement : les hommes ayant le statut le moins élevé, ne pouvant trouver de conjointe plus bas qu'eux sur l'échelle sociale, de même que les femmes ayant le statut le plus élevé, ne pouvant trouver de conjoint plus haut qu'elles, n'ont d'autre choix que le célibat. Ce modèle a aussi mis en lumière certains détails que notre cadre théorique n'avait pas anticipés : l'interaction entre identités professionnelles et rôles de genre, par laquelle les hommes employés non qualifiés prennent bien plus souvent pour conjointes des femmes de statut plus élevé qu'eux que les ouvriers non qualifiés. Il a finalement permis d'identifier des asymétries genrées entre les individus d'origines en col blanc, en col bleu et agricoles. Les évolutions temporelles déjà décrites à l'aide d'une définition *a priori* de l'hypergamie se trouvent pleinement confirmées par ce modèle.

Ainsi, à l'instar du développement des études féminines, l'évolution observée semble être à la fois le signe et un facteur potentiel d'une remise en cause des rôles genrés au sein du couple (Duru-Bellat, 1990, Conclusion). Ce phénomène pourrait jeter les bases d'une évolution de long terme *via* la socialisation des nouvelles générations au sein de familles dans lesquelles la femme est la plus diplômée.

Plus encore, la rapidité des changements observés paraît impliquer que ces normes ne leur ont opposé aucune inertie. Notre étude confirme, dans le cas de la France, la conclusion obtenue en comparant différents pays à différentes dates, selon laquelle dans aucune société on n'observe une préférence relative pour l'hypergamie féminine allant à l'encontre de la structure par diplôme des populations des hommes et des femmes (Esteve, García-Román et Permanyer, 2011). Un tel résultat tend à remettre en cause l'existence même d'une véritable norme d'hypergamie féminine, du moins du point de vue du diplôme. Plus généralement, il pose la question de la possibilité pour un système culturel de résister à une transformation profonde des conditions structurelles dans lesquelles il a vu le jour : sans appui sur les situations concrètes, une norme peut-elle conserver son sens et sa prégnance ? On retrouve

ici le double sens, positif – ce qui est le plus fréquent – et normatif – ce qui devrait être –, du terme *normal*, le premier semblant déterminer le second¹⁸.

Cependant, l'évolution spectaculaire observée ici doit être relativisée. En effet, les conséquences du développement de l'hypogamie féminine ne sont pas absolument évidentes (Guichard-Claudic, Testenoire et Trancart, 2009). On sait en effet que l'augmentation du niveau d'éducation des femmes n'a pas permis de leur garantir des carrières égales à celle des hommes, du fait du maintien de la ségrégation genrée à la fois selon le domaine d'études au sein de chaque niveau (Caille, Lemaire et Vrolant, 2002 ; Duru-Bellat, 2004) et sur le marché du travail (Couppié et Epiphane, 2006). Elle n'a pas non plus permis de modifier profondément la répartition du travail domestique. C'est que le statut d'étudiante ou de diplômée ne suffit pas en lui-même à offrir une position de supériorité au sein du couple, ni une plus grande facilité à renégocier les rôles (Lefevre, 2008) : seule l'activité professionnelle semble avoir un certain effet en ce sens (Zarca, 1990 ; Bauer, 2007 ; Champagne, Pailhé et Solaz, 2014, p. 12-13), du moins lorsque les revenus de la femme sont assez élevés par rapport à ceux de l'homme (Nicole-Drancourt, 1989 ; Ponthieux et Schreiber, 2006). Or, si l'hypergamie relative de diplôme s'est clairement inversée, celle de classe sociale n'a pas complètement disparu, et – surtout – celle de salaire est restée stable parmi les couples bi-actifs. L'hypothèse selon laquelle les rôles respectifs de pourvoyeur de ressources et de responsable de la tenue du ménage sont plus fondamentaux dans la définition des identités genrées que les niveaux relatifs de capital culturel des conjoints ou même que leurs statuts professionnels est confirmée. Ceci expliquerait alors la résistance des rôles genrés au sein du couple malgré l'inversion de l'hypergamie de diplôme – qui serait donc en partie un mirage tant qu'elle ne se traduit pas en espèces sonnantes et trébuchantes. Cette réserve ne doit pas conduire à sous-estimer l'ampleur des évolutions qui ont pris place au cours des quarante dernières années, ni le rythme auquel elles se poursuivent.

18 Soit dans les termes de Pierre Bourdieu (1998, p. 102), « *la statistique spontanée, à travers laquelle se forme la représentation que chacun de nous a du normal* ».

CHAPITRE 5

LES MÉCANISMES DU CÉLIBAT DÉFINITIF DES DIPLÔMÉES : LA FIN DE LA SAINTE-CATHERINE (COHORTES 1920-1970)

L'affaiblissement notable de l'hypergamie féminine de diplôme, de classe et d'origine sociales observé au chapitre précédent s'est accompagné d'une diminution des différences genrées dans la distribution sociale du célibat¹. Ces deux évolutions ont été les plus marquées du point de vue du diplôme, qui a connu les plus fortes transformations structurelles en termes de différences hommes-femmes. Cette concomitance des évolutions de l'hypergamie et du célibat soutient notre hypothèse selon laquelle ces deux phénomènes sont fondamentalement liés. Ce chapitre vise à confirmer cette relation, puis à identifier plus précisément les mécanismes qui en sont à l'origine. Pour cela, nous nous tournons vers des données longitudinales permettant de suivre le processus de première mise en couple des individus : nous ne chercherons donc plus à étudier le célibat à un instant donné, mais le célibat définitif – soit le fait de ne jamais avoir vécu en couple. Ce changement d'approche sera aussi l'occasion de préciser la chronologie des évolutions déjà relevées, en les détaillant par cohorte de naissance, alors que ces dernières étaient jusqu'ici confondues chaque année dans un même échantillon transversal.

Nous nous concentrons ici sur l'hypergamie de diplôme. En effet, l'éducation est le principal déterminant du style de vie et des fréquentations des individus à l'âge de leur mise en couple, mais aussi l'un des indicateurs les plus forts de leur carrière et de leur statut social futurs. À ce titre, elle représente l'une des informations les plus importantes dont disposent – consciemment ou inconsciemment – les individus concernant les conjoints potentiels qu'ils rencontrent. La scolarisation influence aussi l'âge de mise en couple, ce qui

1 Des éléments de ce chapitre (ainsi que du précédent) ont donné lieu à un article à paraître dans la revue *Population* sous le titre « Plus diplômées, moins célibataires : l'inversion de l'hypergamie féminine d'éducation au fil des cohortes en France ».

a des effets non négligeables sur le célibat définitif : selon l'hypothèse proposée par François de Singly (1987, chap. 8) déjà présentée au chapitre précédent, le surcroît de célibat des femmes les plus diplômées s'explique par un décalage entre le calendrier des études et celui de la mise en couple.

C'est cette hypothèse que nous mettrons à l'épreuve dans ce chapitre à travers l'étude de la distribution des âges de première mise en couple selon le niveau de diplôme. Si nos résultats soutiennent cette explication pour les cohortes les plus anciennes, elle ne semble être valable que dans un contexte de prévalence de l'hypergamie féminine et d'une forte incompatibilité entre études et vie en couple. L'inversion de la première, et l'affaiblissement de la seconde du fait du développement de la cohabitation hors mariage ont fait disparaître, parmi les femmes des cohortes récentes, les différences de taux de célibat définitif selon le diplôme. Ainsi, nous montrerons que la limite symbolique des 25 ans, âge à partir duquel les femmes encore célibataires défilaient traditionnellement le jour de la Sainte-Catherine (25 novembre) de chaque année en portant un chapeau caractéristique (Bologne, 2004, p. 133, 412), marquait bel et bien le début d'une nette diminution des chances de première mise en couple qui frappait en particulier les plus diplômées. Mais cette difficulté à se mettre en couple au-delà de 25 ans s'est fortement atténuée en même temps que la tradition des catherinettes tombait en désuétude, participant à la disparition du surcroît de célibat de ce groupe.

Nous commençons par présenter les données des enquêtes Étude de l'histoire familiale (Insee-Ined, 1999) et Biographies et entourage (Ined, 2001), utilisées de manière complémentaire dans ce chapitre, l'une pour le nombre de cohortes qu'elle couvre et son important échantillon, l'autre pour le détail des informations qu'elle fournit. Confirmant les résultats du chapitre précédent, nous montrons dans un premier temps qu'au fil des cohortes l'hypergamie s'est inversée, à la fois en termes de répartition brute des couples, mais aussi en termes relatifs, en recourant à un modèle log-linéaire contrôlant les évolutions structurelles de la population.

Nous présentons dans un second temps un modèle à temps de sortie accéléré fondé sur la distribution log-logistique avec immunité, qui permet d'étudier à la fois la distribution des âges de mise en couple et la proportion de célibataires définitifs au fil des cohortes. Ce modèle met en évidence une résorption complète des écarts dans la proportion de femmes

qui ne vivront jamais en couple cohabitant entre moins diplômées et plus diplômées ; cependant, nuanciant les résultats du chapitre précédent, le célibat définitif des hommes non diplômés reste toujours plus élevé que celui des femmes du même groupe. L'examen de l'évolution avec l'âge des chances de mise en couple montre que le célibat définitif des femmes les plus diplômées s'expliquait par une difficulté à se mettre en couple après 25 ans, qui nous paraît trouver sa source dans la prévalence de l'hypergamie féminine, ainsi que dans l'incompatibilité entre études et vie en couple – deux facteurs qui ne jouent plus parmi les cohortes récentes.

Néanmoins, cette évolution ne s'est pas accompagnée d'une remise en cause des différences entre hommes et femmes aussi nette du point de vue du célibat définitif et des âges de mise en couple que de celui du célibat à un instant donné mesuré au chapitre précédent.

1 Les données : Étude de l'histoire familiale et Biographies et entourage

Si nous avons jusqu'ici mesuré l'hypergamie à partir d'enquêtes transversales portant sur les ménages (les enquêtes Emploi), les enquêtes permettant d'étudier des évolutions au fil des cohortes à travers la première mise en couple sont malheureusement rares. Nous mobilisons ici deux enquêtes complémentaires². D'une part, l'enquête Étude de l'histoire familiale (Insee-Ined, 1999, voir Lefèvre et Filhon, 2005) est la plus récente des enquêtes Famille à contenir des informations concernant l'âge de première mise en couple³. Adossée au recensement de la population, elle offre l'avantage d'un très important échantillon de 145 000 hommes et 235 000 femmes, permettant d'étudier finement les évolutions au fil des cohortes. Nous restreignons ici notre analyse aux individus nés entre 1920 et 1970, qui avaient entre 29 et 79 ans à la date de l'enquête : d'un côté, il s'agit de limiter les biais dus à la mortalité différentielle (Blanpain, 2011) ; de l'autre, de ne pas abusivement tenter de

2 Les fichiers de l'enquête Étude de l'histoire familiale nous ont été mis à disposition par l'ADISP (Centre Maurice Halbwachs) *via* le réseau Quételet. Ceux de Biographies et entourage nous ont été rendus accessibles par notre association à l'Unité 6 (Mobilité, logement et entourage) de l'Ined. Que ces deux institutions, et les personnes responsables de ces enquêtes, en soient ici remerciées.

3 En effet, de même que la plupart des enquêtes sur les ménages, l'enquête Famille et logements 2011 ne contient d'informations qu'à propos des unions en cours. L'âge de première mise en couple n'est donc disponible que pour les individus n'ayant jamais connu de séparation.

décrire une cohorte à partir des individus s'étant mis en couple particulièrement tôt. Dans les deux cas, les biais ne peuvent être véritablement écartés concernant les cohortes extrêmes : celles-ci sont incluses à titre indicatif uniquement, et les conclusions, pour être robustes, doivent plutôt se fonder sur les cohortes nées entre 1930 et 1960 (âgées de 39 à 69 ans à la date de l'enquête).

Afin de situer les cohortes étudiées ici par rapport à l'évolution mesurée selon l'année d'enquête dans les chapitres précédents, notons que la cohorte 1920 était déjà présente dans notre échantillon issu des enquêtes Emploi en 1969 (à l'âge de 39 ans) et en est sortie en 1980 (à l'âge de 60 ans), alors que la cohorte 1970 est entrée dans notre échantillon en 2000 (à l'âge de 30 ans) et y était encore présente en 2011 (à l'âge de 41 ans). Ce sont donc les mêmes évolutions que nous observons, en prenant deux points de vue différents.

Les données présentent cependant une limite de taille : le diplôme du conjoint étant retrouvé à partir de la fiche individuelle de recensement, il n'est disponible que pour les couples cohabitant au moment de l'enquête. Les premiers couples des individus ayant connu une séparation (éventuellement suivie d'une remise en couple) ne peuvent donc être étudiés. Ainsi, on peut craindre un léger biais si les couples s'étant séparés étaient différents de l'ensemble des premiers couples, par exemple s'il s'agissait en priorité de couples dans lesquels l'homme est moins diplômé que sa conjointe, le modèle traditionnel de la famille excluant à la fois l'hypogamie féminine et la séparation ou le divorce. Ce biais est particulièrement problématique pour la mesure des évolutions temporelles, puisque la proportion d'individus ayant vécu avec un seul conjoint varie au fil des cohortes (Tableau 5.1) : elle diminue légèrement pour les cohortes nées après guerre (effet de génération). La proportion d'individus vivant toujours avec leur premier conjoint suit la même tendance chez les hommes, mais elle augmente au contraire chez les femmes (effet d'âge lié à la mortalité plus forte de leurs conjoints). L'ampleur de cet effet est vraisemblablement relativement limitée du fait de la faible proportion de couples concernée – en particulier si l'on exclut les cohortes extrêmes.

Tableau 5.1 : Répartition des individus selon leur histoire conjugale et les données disponibles par cohorte et par sexe

	Effectif	Ensemble	Ayant vécu en couple	N'a eu qu'un conjoint	Toujours avec son premier conjoint	Sans valeurs manquantes
Étude de l'histoire familiale : Femmes						
1920-29	21 523	100	92	84	45	36
1930-39	26 238	100	93	85	61	50
1940-49	29 815	100	94	83	66	56
1950-59	38 078	100	93	79	65	57
1960-69	39 410	100	90	76	64	57
Étude de l'histoire familiale : Hommes						
1920-29	10 812	100	92	85	73	61
1930-39	15 544	100	92	84	74	63
1940-49	19 529	100	93	81	71	63
1950-59	24 791	100	90	77	67	59
1960-69	25 720	100	82	73	62	56
Biographies et entourage : Femmes						
1930-39	489	100	91	68	52	50
1940-50	692	100	94	72	46	43
Biographies et entourage : Hommes						
1930-39	408	100	95	75	55	54
1940-50	583	100	95	71	49	47

Source : enquêtes Étude de l'histoire familiale (Insee-Ined, 1999) et Biographies et entourage (Ined, 2001)

Champ : individus nés en France

« Sans valeurs manquantes » indique que les diplômés de l'individu et de son premier conjoint, la date de mise en couple et la date de naissance de l'individu sont renseignés. Les analyses qui suivent ne mobilisent jamais toutes ces informations ensemble : cette colonne indique un seuil minimum théorique concernant les échantillons retenus.

D'autre part, on peut craindre que le diplôme à la date de l'enquête ne reflète pas parfaitement le diplôme des conjoints aux alentours de la mise en couple, c'est-à-dire à la fin de la formation initiale, du fait de la possibilité de recourir à la formation continue ; celle-ci

reste cependant très peu fréquente en France. On sait aussi qu'existe un effet de déclaration variable avec l'âge (Baudelot, 1989), mais son ampleur est cependant limitée. On peut noter que l'étude de l'homogamie de diplôme réalisée par Mélanie Vanderschelden (2006a) à partir des mêmes données concorde avec celle que nous avons réalisée au chapitre 2 en identifiant une tendance à la baisse⁴ : le biais introduit n'est donc pas rédhibitoire. Néanmoins, afin de se prémunir contre cette difficulté, il importe de ne pas accorder trop de crédit aux évolutions temporelles les plus faibles : nous ne commenterons donc que les tendances les plus nettes.

Les lacunes de l'enquête Étude de l'histoire familiale nous ont amené à recourir à une deuxième enquête, Biographies et entourage (Ined, 2001, voir Bonvalet et Lelièvre, 2012), qui fournit des informations détaillées sur le parcours conjugal de la personne interrogée et de ses conjoints successifs, répondant ainsi aux difficultés que nous venons de soulever⁵. Cette enquête, plus petite (2830 individus), ne permet cependant pas toujours de mettre en évidence des résultats statistiquement significatifs. Portant sur les personnes nées entre 1930 et 1950 et résidant en Île-de-France en 2001, elle nous servira de confirmation pour la description des cohortes nées avant-guerre et dans l'immédiat après-guerre. Néanmoins, puisque les champs géographiques des deux enquêtes diffèrent, on ne doit pas s'attendre à obtenir des résultats parfaitement identiques : les proportions d'individus ayant vécu avec un seul conjoint, bien moins élevées en Île-de-France, en fournissent une illustration (Tableau 5.1).

Nous nous intéressons ici à la première cohabitation (qu'elle soit associée ou non à un mariage), l'étude du premier mariage ne paraissant pas appropriée tant le sens de cet événement a changé sur la période étudiée. Si le sens de la cohabitation a aussi évolué au fil des cohortes, ce type d'union n'étant en particulier plus un simple prélude au mariage et les séparations étant plus fréquentes (Toulemon, 1996), cette définition semble cependant la plus adaptée puisqu'il s'agit ici d'étudier le premier engagement conjugal. Les premiers couples cohabitants apparaissent comme relativement stables : les « unions éphémères » durant moins de 3 ans ne constituent dans la période étudiée qu'une part limitée, inférieure à 20 %, des premiers couples cohabitants (Villeneuve-Gokalp, 1990). En outre, selon les

4 Cette étude n'a en revanche pas identifié la tendance à la baisse de l'homogamie socioprofessionnelle, la profession évoluant beaucoup plus nettement que le diplôme avec l'âge.

5 Les deux enquêtes utilisées ici ont déjà été utilisées conjointement pour une étude des manières de déclarer son histoire conjugale et familiale (Lelièvre et Mazuy, 2005).

définitions retenues par les enquêtes utilisées ici, la vie commune doit avoir duré plus de six mois pour Étude de l'histoire familiale, et plus d'un an pour Biographies et entourage⁶. Le choix de s'en tenir aux premières mises en couple vise à permettre la modélisation du processus de mise en couple et de ses évolutions par cohortes, de manière à analyser le fonctionnement du marché conjugal, ce qui ne serait pas possible en confondant des unions de rangs différents.

Nous recourons à une classification de diplôme relativement détaillée, de manière à décrire des distinctions qui font sens aussi bien au sein des cohortes les plus anciennes qu'au sein des plus récentes. Les catégories retenues sont : aucun diplôme ; certificat d'études primaires (CEP) ; CAP, BEP, BEPC ; baccalauréat général, technologique ou professionnel ; diplôme du premier cycle du supérieur (inférieur à la licence) ; diplôme des second ou troisième cycles du supérieur. Notons enfin que le diplôme considéré est celui finalement obtenu par les individus, que ce soit avant ou après la mise en couple. Ce choix, inévitable du fait de la structure des données, revient à décrire le résultat de la mise en couple, plutôt que ses déterminants. En tout état de cause, la mise en couple ayant très rarement lieu longtemps avant la fin des études, on peut raisonnablement considérer que les conjoints avaient à ce moment une idée assez précise de leur niveau d'éducation final.

Notons finalement que toutes les analyses utilisent les pondérations fournies par les enquêtes, et que l'âge de mise en couple est mesuré en mois.

2 L'inversion de l'hypergamie de diplôme au fil des cohortes

Les positions relatives des hommes et des femmes sur le marché conjugal du point de vue du diplôme se sont radicalement transformées entre la cohorte née en 1920 et celle née en 1960. Afin de confirmer les résultats du chapitre précédent et de préciser les évolutions du point de vue des cohortes plutôt que des années d'enquête, nous commençons par décrire l'évolution brute de la composition des couples en termes de diplômes des conjoints, avant de recourir à un modèle log-linéaire pour mesurer celle de l'hypergamie relative.

⁶ Cette dernière enquête ne précise pas si une cohabitation doit avoir eu lieu, ni sur quelle durée. Néanmoins, la date de mise en couple renseignée est celle de la mise en ménage.

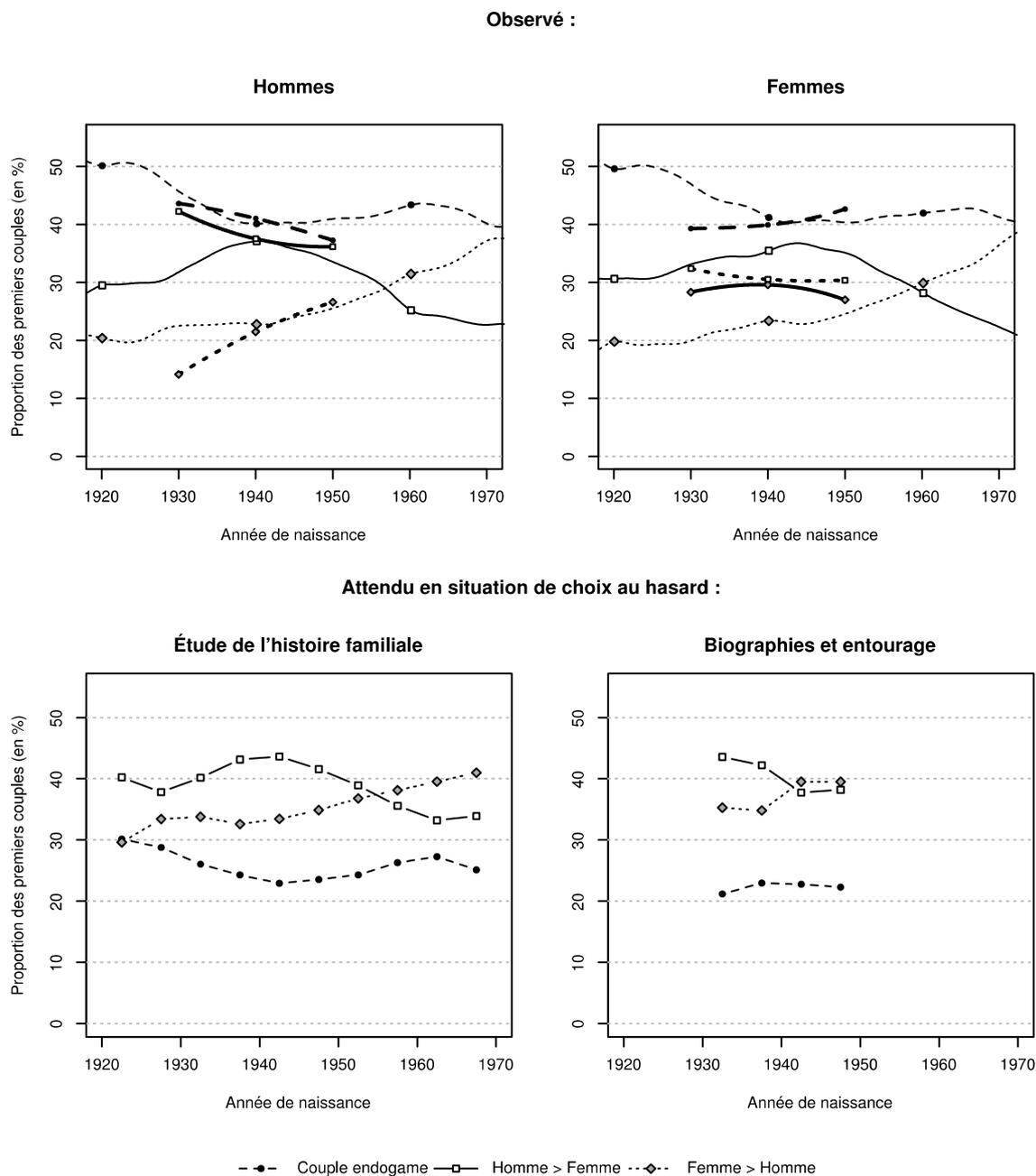
2.1 L'hygogamie féminine est devenue la situation la plus fréquente

L'évolution de l'hypergamie absolue (Figure 5.1, partie supérieure) est cohérente avec nos observations précédentes, mais fait ressortir des détails qui n'étaient pas apparus dans l'analyse transversale.

Nous pouvons tout d'abord noter que les situations respectives des hommes et des femmes au sein de leurs premiers couples, qui ne se recouvrent pourtant pas complètement du fait que certaines unions ne sont un premier couple que pour l'un des deux conjoints, sont presque parfaitement symétriques. Chez les deux sexes, il est clair que les couples dans lesquels l'homme est plus diplômé que sa conjointe (hypergamie féminine ou hypogamie masculine) étaient loin de représenter la majorité des couples, même parmi les cohortes d'avant-guerre (Guichard-Claudic, Testenoire et Trancart, 2009) : ce sont les couples endogames qui étaient de loin les plus courants (50 % dans notre classification pour la cohorte 1920). L'asymétrie entre hommes et femmes ne s'exprime donc qu'au sein des situations d'exogamie, par la différence entre les proportions de couples dans lesquels l'homme est le plus diplômé (30 % pour la même cohorte) et ceux dans lesquels c'est la femme qui l'est (20 %).

On peut distinguer deux phases de l'évolution au fil des cohortes. Dans une première phase, correspondant aux cohortes nées dans les années 1920 et 1930, la proportion de couples endogames diminue en faveur des deux types de couples exogames (hypergamie et hypogamie), ce qui correspond en termes de cohortes à l'évolution mise en évidence au chapitre 1. Cette phase ne correspond pas à une nette évolution des différences genrées, même si on peut souligner qu'elle voit l'hypergamie féminine se développer un peu plus rapidement que la situation inverse.

Figure 5.1 : Relation entre les diplômes des conjoints au sein des premiers couples selon l'année de naissance (proportions observées et proportions attendues sous l'hypothèse de choix aléatoire du conjoint)



Dans la partie supérieure de la figure, les courbes fines correspondent à l'enquête Étude de l'histoire familiale, celles épaisses à Biographies et entourage. Pour améliorer la lisibilité, une régression locale de degré 1 (lissage non paramétrique) est utilisée pour les pourcentages observés (cf. la légende de la figure 3.3, p. 202), alors que les pourcentages attendus sont calculés sur la base de cohortes quinquennales. En effet, dans ce dernier cas il est nécessaire de faire l'hypothèse qu'il existe un groupe de référence (ici, la cohorte

quinquennale) au sein duquel les individus trouvent approximativement leur conjoint : une approche continue n'est pas possible.

Les premiers couples observés des hommes et des femmes ne sont pas exactement identiques puisqu'une personne peut se mettre en couple pour la première fois avec quelqu'un ayant déjà vécu en couple. La situation théorique de mise en couple au hasard que nous retenons ici ne tient pas compte de cette possibilité, ni de celle du célibat, ce qui se justifie dans cette section par le fait que nous nous intéressons uniquement à l'effet de la structure par diplôme d'une cohorte de naissance.

Champ : premiers couples des individus nés en France et n'ayant pas connu de séparation pour Étude de l'histoire familiale ; premiers couples des individus nés en France et résidant en Île-de-France en 2001 pour Biographies et entourage

Source : enquêtes Étude de l'histoire familiale (Insee-Ined, 1999) et Biographies et entourage (Ined, 2001)

Dans une seconde phase, entamée avec les cohortes 1940-1945⁷, la proportion de couples dans lesquels la femme est la moins diplômée commence à chuter rapidement, alors que la proportion de couples dans lesquels elle est la plus diplômée augmente au contraire à un rythme désormais soutenu. Dès les cohortes nées à la fin des années 1950, les cas d'hyogamie féminine dépassent ceux d'hypergamie. Cette évolution aboutit à ce qu'au sein des cohortes nées au début des années 1970, les premiers couples dans lesquels la femme possède un diplôme supérieur à celui de son conjoint sont aussi fréquents que les couples endogames (40 % des couples). La tendance semble en outre devoir se prolonger pour les cohortes plus récentes. Cette chronologie de l'évolution de l'hypergamie concorde pleinement avec celle des nombreuses transformations ayant affecté la famille sur la même période. Dans leur histoire des *baby-boomers*, Catherine Bonvalet, Céline Clément et Jim Ogg (2011, p. 118) soulignent en effet que ces derniers se sont « engouffrés » dans une brèche ouverte par les générations nées pendant la Seconde Guerre mondiale : l'image s'applique parfaitement ici.

Une illustration frappante du renversement opéré autour de ces générations charnières est fournie par l'évolution de l'hypergamie des diplômés intermédiaires. Parmi les titulaires d'un diplôme du secondaire inférieur (CAP/BEP/BEPC) nés dans les années 1920, 27 % des femmes, mais seulement 7 % des hommes, avaient pris pour premier conjoint une personne

7 Le point d'inflexion semble arriver plus tôt chez les hommes que chez les femmes, ce qui est cohérent avec le fait que les hommes de ces cohortes sont en moyenne plus âgés que leurs conjointes d'environ 3 ans (Vanderschelden, 2006b).

plus diplômée. Ces chiffres se sont nettement inversés, passant à respectivement 20 % et 34 % pour les générations nées dans les années 1960.

Les données de Biographies et entourage, qui portent sur une aire géographique plus restreinte (l'Île-de-France) différente du reste de la France du point de vue du niveau d'éducation, confirment ces tendances temporelles, quoique moins clairement chez les premiers couples des femmes que chez ceux des hommes. Le manque de symétrie peut sans doute être attribué au fait que les premiers conjoints des interrogés ne vivent pas toujours en Île-de-France (séparation suivie d'une migration par exemple), et que les mécanismes de migration ne sont pas nécessairement les mêmes chez les deux sexes (migration de célibataires plutôt que de couples). Il apparaît que les femmes nées entre 1930 et 1950 et résidant en Île-de-France en 2001 ont connu une situation plus égalitaire dans leur premier couple que les femmes des mêmes cohortes vivant dans le reste de la France.

Les évolutions que nous venons de commenter sont pour l'essentiel une conséquence directe de l'élévation du niveau d'éducation des femmes : dès les cohortes nées au milieu des années 1950, les femmes sont globalement plus diplômées que les hommes⁸. En effet, la répartition des premiers couples qui serait attendue si les individus choisissaient leur conjoint au hasard parmi des membres de leur cohorte quinquennale de naissance du sexe opposé suit globalement les mêmes tendances (Figure 5.1, partie inférieure). La différence la plus notable entre la situation observée et la situation théorique tient au fait que, de manière tout à fait classique, les couples endogames sont sur-représentés, et les deux autres types de couples sous-représentés, parmi les couples observés dans toutes les cohortes. On peut aussi remarquer que les évolutions de la situation observée sont plus marquées que celles de la situation théorique : comme nous l'avons déjà noté au chapitre précédent, la massification scolaire n'explique pas l'intégralité de l'évolution de l'hypergamie. L'utilisation d'un modèle log-linéaire va nous permettre d'observer plus finement si les tendances respectives à l'hypergamie des hommes et des femmes ont évolué au fil des cohortes, au-delà de l'effet impliqué par ces évolutions structurelles.

8 La probabilité qu'en choisissant au hasard un homme et une femme au sein d'une cohorte, la femme soit la plus diplômée des deux (Esteve, García-Román et Permanyer, 2012, p. 540) dépasse 1/2 pour la première fois pour la cohorte 1955. Cet indicateur passe de 0,44 pour les cohortes nées dans les années 1920 à 0,54 pour celles nées dans les années 1960.

2.2 L'inversion persiste au-delà des évolutions structurelles

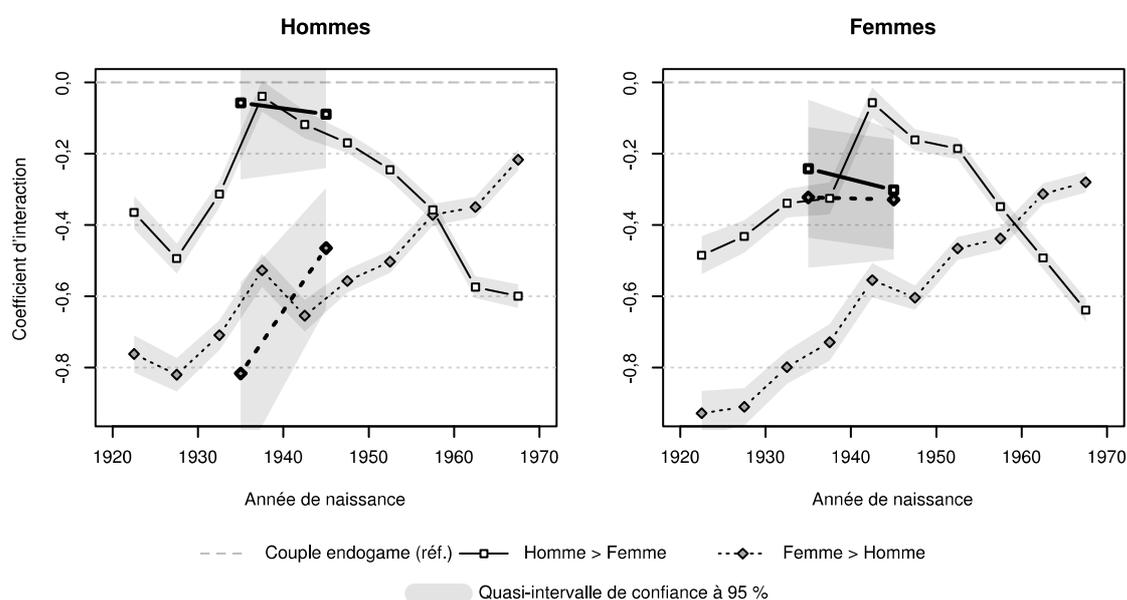
Les tendances absolues observées dans la section précédente persistent lorsque l'on contrôle les évolutions de la structure par diplôme des populations masculine et féminine à l'aide d'un modèle log-linéaire (Figure 5.2). Les mêmes phases peuvent être distinguées, un retournement de tendance concernant l'hypergamie féminine se produit autour de la cohorte 1940-1945 ; les situations dans lesquelles la femme est la plus diplômée font jeu égal avec celles dans lesquelles c'est l'homme qui l'est à partir de la cohorte 1955-1959. Dans les limites autorisées par les intervalles de confiance⁹, les deux cohortes décennales de Biographies et entourage confirment ces tendances chez les hommes, mais moins nettement chez les femmes, du fait encore une fois d'une hypogamie bien plus élevée que dans le reste de la France.

Ces résultats indiquent que le choix du conjoint suit et dépasse les transformations de la structure par diplôme de la population : au sein des générations nées dans les années 1960, les couples dans lesquels la femme est moins diplômée que son conjoint sont moins fréquents que les cas inverses par rapport à ce que l'on attendrait si la mise en couple s'effectuait au hasard. Les résultats du chapitre précédent sont donc pleinement confirmés par l'analyse en termes de cohortes. En particulier, on observe ici aussi une synchronisation parfaite entre hypergamie absolue et hypergamie relative, suivant le résultat déjà établi par Albert Esteve, Joan García-Román et Iñaki Permanyer (2011) à propos de nombreux pays. La norme d'hypergamie féminine ne semble pas avoir opposé la moindre inertie à l'évolution de la structure par diplôme de la population.

Si elle est aussi profonde qu'il y paraît, cette évolution concernant l'hypergamie des personnes en couple doit s'être accompagnée d'une transformation elle aussi notable des chances de vie en couple des individus qui étaient les moins favorisés sur le marché conjugal : les hommes non diplômés, et les femmes diplômées. Il s'agit donc maintenant d'analyser la mise en couple des individus – et son absence – par groupe de diplôme, à travers le célibat définitif et l'âge de première mise en couple.

⁹ Un modèle postulant que l'hypergamie et l'hypogamie sont égales pour les deux cohortes décennales décrit moins bien les données que le modèle retenu à la probabilité critique de 6 % chez les hommes, mais cette probabilité est de 90 % chez les femmes.

Figure 5.2 : Hypergamie et hypogamie relatives de diplôme au sein des premiers couples selon l'année de naissance dans un modèle log-linéaire



Les courbes fines correspondent à l'enquête Étude de l'histoire familiale, celles épaisses à Biographies et entourage.

Ce modèle prend pour référence une situation dans laquelle les individus prendraient leur conjoint au hasard parmi tous les conjoints des membres de leur cohorte quinquennale de naissance (les célibataires étant ainsi considérés comme hors de ce marché conjugal). Son équation s'écrit :

$$\log m_{hfc} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_c^C + \lambda_{hc}^{HC} + \lambda_{fc}^{FC} + \lambda_c^{C(H>F)} \mathbb{1}_{h>f} + \lambda_c^{C(H<F)} \mathbb{1}_{h<f}$$

avec m_{hfc} les effectifs modélisés pour le diplôme de l'homme h et celui de la femme f , $\mathbb{1}_{h>f}$ et $\mathbb{1}_{h<f}$ les indicatrices correspondant respectivement aux situations d'hypergamie et d'hypogamie féminine, et $\lambda^{C(H>F)}$ et $\lambda^{C(H<F)}$ les coefficients (repris sur la figure) mesurant la sur- ou sous-représentation de ces situations par rapport aux configurations endogames dans la cohorte c . L'estimation est réalisée simultanément pour les hommes et les femmes (représentant chacun une couche séparée), en introduisant des interactions entre le sexe et tous les paramètres décrits ci-dessus. Le modèle concernant Étude de l'histoire familiale est fondé sur 329 007 observations et présente une déviance de 806 913 pour 1 797 degrés de liberté. Celui concernant Biographies et entourage est fondé sur 2 172 observations et présente une déviance de 5 701 pour 360 degrés de liberté.

Un quasi-intervalle de confiance (Firth et Menezes, 2004) permet de comparer l'ensemble des points entre eux, et pas seulement chaque point à la référence (0 pour endogame).

Champ : premiers couples des individus nés en France et n'ayant pas connu de séparation pour Étude de l'histoire familiale ; premiers couples des individus nés en France et résidant en Île-de-France en 2001 pour Biographies et entourage

Source : enquêtes Étude de l'histoire familiale (Insee-Ined, 1999) et Biographies et entourage (Ined, 2001)

3 Modéliser la première mise en couple

3.1 Le modèle log-logistique avec immunité

L'étude du célibat dit définitif, autrement dit le fait de ne s'être jamais mis en couple, est relativement délicate puisque le phénomène ne peut être parfaitement observé qu'à la fin de la vie des intéressés. L'analyse des évolutions temporelles s'en trouve donc nettement retardée. Mesurer la part de personnes ayant vécu en couple à un âge donné permet de limiter cette difficulté, mais ne permet pas de tenir compte des évolutions de l'âge de première mise en couple : un retard de la mise en couple se traduit ainsi par une augmentation du taux de célibat observé à un âge donné, augmentation qui pourra cependant être compensée par une plus grande propension à se mettre en couple à des âges élevés. Ce type de phénomène peut typiquement s'observer suite à une augmentation de l'âge de fin d'études au fil des cohortes.

Afin de distinguer le taux de célibat définitif de cet effet d'âge, nous recourons ici à un modèle log-logistique à temps de sortie accéléré, qui permet de prédire avec un degré de confiance assez élevé la proportion d'individus qui n'auront jamais vécu en couple au cours de leur vie, à partir de l'observation d'une cohorte jusqu'à un âge donné. Les premières mises en couple étant nettement ralenties au-delà de 35 ans, il est possible de fournir une estimation fiable de la proportion de célibataires définitifs des cohortes nées dans les années 1960, voire d'esquisser des tendances concernant les générations plus récentes.

Nous utilisons ici le modèle log-logistique avec immunité¹⁰ (Brüderl et Diekmann, 1995) qui a précisément été élaboré pour l'étude de la mise en couple. Ce modèle présente l'intérêt de décrire très précisément les rythmes de mise en couple réellement observés, et d'offrir une interprétation intéressante en termes de processus de diffusion au sein de chaque groupe social (Diekmann, 1989 ; voir Todd, Billari et Simão, 2005 pour un approfondissement de cette approche). Il comprend trois paramètres, estimés séparément pour chaque groupe de diplôme : la proportion de célibataires définitifs ; l'âge médian de mise en couple des individus qui vivront au moins une fois en couple ; la forme de la

10 « Immunité » désignant ici le fait de ne jamais connaître l'événement d'intérêt ; ce type de modèle est aussi parfois qualifié de *movers-stayers*. En l'absence d'un terme à la fois neutre et approprié à la question de la mise en couple, nous retenons l'idée d'immunité.

distribution des âges de mise en couple (*cf.* Encadré 5.1)¹¹. Au sein de chaque groupe de diplôme, ces paramètres sont autorisés à varier au fil des cohortes grâce à des splines cubiques restreintes de degré 3, qui permettent d'étudier des évolutions temporelles de manière flexible (*cf.* Encadré 5.2). En effet, on sait que l'évolution de l'âge de mise en couple n'a pas été linéaire, mais en « U », les *baby-boomers* ayant connu les âges médians de mise en couple les plus bas du siècle (Prioux, 2003) – ce que nous retrouvons ici. L'extrapolation de la proportion de célibataires définitifs au sein des cohortes les plus récentes réalisée ici repose ainsi d'une part sur le comportement de ces cohortes observé jusqu'à la date de l'enquête, et d'autre part sur la prolongation grâce aux splines des distributions observées pour les cohortes les plus anciennes.

Le modèle est estimé dans l'environnement R (R Core Team, 2014) grâce à une addition intégrée au paquet *flexsurv* (Jackson, 2014)¹². L'âge de mise en couple est mesuré en mois, à partir de l'année et du mois de naissance et de ceux de mise en couple.

Dans la suite du chapitre, nous mobilisons uniquement l'enquête Étude de l'histoire familiale. En effet, puisque le diplôme du conjoint n'est pas utile dans cette analyse, le champ peut être étendu à l'ensemble des premiers couples, y compris ceux s'étant séparés à la date de l'enquête. Il ne semble donc pas nécessaire de surcharger la présentation en utilisant deux enquêtes en parallèle.

Encadré 5.1 : Équations et estimation de la distribution log-logistique avec immunité

La distribution log-logistique avec immunité, introduite par Josef Brüderl et Andreas Diekmann (1995, p. 166), est une généralisation de la distribution log-logistique. Nous présentons ici la paramétrisation retenue par les auteurs (utilisant λ , p et k), puis une paramétrisation plus standard (utilisant $\alpha = 1/\lambda$, $\beta = p$ et k) dans laquelle α est égal au temps médian d'occurrence de l'événement (pour les individus qui le connaîtront), ce qui facilite l'interprétation des résultats. β représente la forme de la distribution : en

11 Ce modèle pourrait être étendu de manière à prendre en compte des variations potentielles concernant l'âge minimal considéré comme socialement acceptable pour se mettre en couple à l'aide du modèle proposé par Francesco Billari (2001, p. 18), qui introduit un paramètre supplémentaire identifiant cet âge limite.

12 Le code permettant d'estimer les modèles utilisés ici est disponible en ligne sur la page personnelle de l'auteur. Nous tenons à remercier Christopher Jackson pour son aide et pour avoir accepté nos modifications.

particulier, lorsque β tend vers $+\infty$, son mode et sa moyenne tendent vers α , alors que sa dispersion, son asymétrie et son aplatissement (moments d'ordre 2 à 4) tendent vers 0. Enfin, k est la proportion d'individus qui connaîtront finalement l'événement ; lorsque $k = 1$, on se ramène à la distribution log-logistique classique. Les valeurs de ces paramètres varient en respectant les contraintes suivantes : $\alpha > 0$, $\beta > 0$ et $0 < k < 1$.

Fonction de répartition :

$$F(t) = k \frac{(\lambda t)^p}{1 + (\lambda t)^p} = k \frac{(t/\alpha)^\beta}{1 + (t/\alpha)^\beta}$$

Fonction de séjour (ou de survie) :

$$S(t) = 1 - F(t) = \frac{1 + (1 - k)(\lambda t)^p}{1 + (\lambda t)^p} = \frac{1 + (1 - k)(t/\alpha)^\beta}{1 + (t/\alpha)^\beta}$$

Fonction de densité de probabilité :

$$f(t) = k \frac{\lambda p (\lambda t)^{p-1}}{(1 + (\lambda t)^p)^2} = k \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{(1 + (t/\alpha)^\beta)^2}$$

Fonction de chances instantanées :

$$h(t) = k \frac{\lambda p (\lambda t)^{p-1}}{[1 + (\lambda t)^p][1 + (1 - k)(\lambda t)^p]} = k \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{[1 + (t/\alpha)^\beta][1 + (1 - k)(t/\alpha)^\beta]}$$

Lorsque le temps écoulé tend vers l'infini, la proportion d'individus ayant fait l'expérience de l'événement tend asymptotiquement vers k :

$$\lim_{t \rightarrow +\infty} F(t) = 1 - \lim_{t \rightarrow +\infty} S(t) = k$$

Cette évolution se reflète dans le fait que les chances instantanées de connaître l'événement tendent vers zéro :

$$\lim_{t \rightarrow +\infty} h(t) = \lim_{t \rightarrow +\infty} f(t) = 0$$

L'estimation par le maximum de vraisemblance du modèle est réalisée de manière classique sur une échelle logarithmique (maximisation de la log-vraisemblance). Par

ailleurs, puisque les paramètres α , β et k ne sont pas autorisés à prendre l'ensemble des valeurs réelles, des transformations sont utilisées pour s'assurer que les valeurs estimées demeurent dans l'intervalle approprié. Ainsi, une transformation logarithmique projette α et β de l'intervalle $]0 ; +\infty[$ vers $] -\infty ; +\infty[$. La fonction exponentielle permet ensuite de réaliser l'opération inverse. Quant à k , la fonction logit = $\log k/(1 - k)$ est utilisée afin de passer de l'intervalle $]0 ; 1[$ à $] -\infty ; +\infty[$, et la fonction logistique $1/(1 + \exp(-x))$ pour se ramener à la valeur d'origine.

Si les logiciels d'optimisation sont généralement capables d'estimer les dérivées des fonctions considérées par dérivation numérique, cette méthode est très coûteuse en calcul lorsque les échantillons en jeu sont de grande taille. Il est donc utile de disposer d'une forme fermée des dérivées nécessaires à la maximisation de la log-vraisemblance du modèle : les dérivées partielles du logarithme de la fonction de survie $S(t)$ et de la fonction de chances instantanées $h(t)$ selon les versions transformées des trois paramètres α , β et k . Ces dérivées ont été calculées à l'aide du logiciel de calcul formel Maxima et de développements manuels. Leur exactitude a été ensuite vérifiée par différentiation numérique.

Dérivées partielles de $\log S(t)$:

$$\frac{\partial \log S(t)}{\partial \log \alpha} = \frac{\partial \alpha}{\partial \log \alpha} \times \frac{\partial \log S(t)}{\partial \alpha} = -\beta \left(\frac{(1-k)(t/\alpha)^\beta}{1 + (1-k)(t/\alpha)^\beta} - \frac{(t/\alpha)^\beta}{1 + (t/\alpha)^\beta} \right)$$

$$\frac{\partial \log S(t)}{\partial \log \beta} = \frac{\partial \beta}{\partial \log \beta} \times \frac{\partial \log S(t)}{\partial \beta} = \beta \log(t/\alpha) \left(\frac{(1-k)(t/\alpha)^\beta}{1 + (1-k)(t/\alpha)^\beta} - \frac{(t/\alpha)^\beta}{1 + (t/\alpha)^\beta} \right)$$

$$\frac{\partial \log S(t)}{\partial \logit k} = \frac{\partial k}{\partial \logit k} \times \frac{\partial \log S(t)}{\partial k} = -k \frac{(1-k)(t/\alpha)^\beta}{1 + (1-k)(t/\alpha)^\beta}$$

Dérivées partielles de $\log h(t)$:

$$\frac{\partial \log h(t)}{\partial \log \alpha} = \frac{\partial \alpha}{\partial \log \alpha} \times \frac{\partial \log h(t)}{\partial \alpha} = -\beta \left(1 - \frac{2(t/\alpha)^\beta}{1 + (t/\alpha)^\beta} \right)$$

$$\frac{\partial \log h(t)}{\partial \log \beta} = \frac{\partial \beta}{\partial \log \beta} \times \frac{\partial \log h(t)}{\partial \beta} = 1 + \beta \log(t/\alpha) \left(1 - \frac{2(t/\alpha)^\beta}{1 + (t/\alpha)^\beta} \right)$$

$$\frac{\partial \log h(t)}{\partial \logit k} = \frac{\partial k}{\partial \logit k} \times \frac{\partial \log h(t)}{\partial k} = 1 - k$$

$$\text{avec } \frac{\partial \alpha}{\partial \log \alpha} = \alpha, \quad \frac{\partial \beta}{\partial \log \beta} = \beta, \quad \text{et } \frac{\partial k}{\partial \logit k} = k(1 - k)$$

Encadré 5.2 : Les splines cubiques restreintes

Les splines cubiques restreintes (Fox, 2000, chap. 6 ; Harrell, 2001, sect. 2.4 ; Keele, 2008, chap. 3) permettent de modéliser un effet non linéaire d'une variable explicative numérique (comme ici l'année de naissance et l'âge) sur la variable dépendante dans tous types de modèles de régression. Le principe des splines consiste à découper l'échelle des valeurs prises par la variable explicative en segments sur chacun desquels l'effet de la variable suit une fonction polynomiale de forme différente, sans rupture en passant d'un segment à l'autre. Ces segments sont généralement définis de manière à diviser l'échelle des valeurs en intervalles de même amplitude, ou encore à inclure un nombre égal d'individus de l'échantillon dans chaque segment (quantiles). Avec des splines linéaires (les plus simples), l'effet de la variable explicative suit ainsi une ligne brisée qui change de pente à chaque nœud reliant deux segments consécutifs, tout en étant partout continue.

Les splines cubiques, plus souples, recourent sur chaque segment à un polynôme de degré 3 pour décrire l'effet. La fonction d'ensemble décrivant l'effet de la variable explicative est continue, de même que ses dérivées de premier et de second ordre, aux nœuds reliant les différents segments. Sa courbe est donc parfaitement lisse, et les points de rupture totalement invisibles : en fait, le choix de la position des nœuds n'a qu'une faible incidence sur les résultats, ce qui rapproche les splines cubiques de la famille des méthodes non paramétriques.

Enfin, les splines cubiques restreintes (dites aussi splines naturelles) sont des splines cubiques que l'on contraint à prendre une forme linéaire au-delà des nœuds extrêmes des deux côtés de l'échelle des valeurs. Cette restriction, qui revient à une régression linéaire classique en dehors des nœuds, permet d'éviter que la relation prenne une forme trop invraisemblable aux extrémités de l'échelle des valeurs, et autorise éventuellement des extrapolations au-delà des données observées.

Au total, cette technique permet donc de modéliser de manière souple une relation non linéaire, sans contraindre *a priori* la forme qu'elle prend. Elle constitue une généralisation de la régression linéaire classique, et une alternative à deux solutions communément utilisées. La première consiste à discrétiser manuellement l'échelle des valeurs en intervalles définis à l'avance (par exemple, des décennies de naissance), et à introduire dans le modèle une série de variables indicatrices, chacune donnant lieu à l'estimation d'un coefficient. Cette technique présente l'inconvénient d'introduire des effets de seuil et de consommer un nombre important de degrés de liberté ; le modèle ne retient en effet aucune information concernant l'ordre des intervalles et les écarts entre leurs bornes. Au contraire, les splines cubiques restreintes respectent la continuité de l'effet de part et d'autre des nœuds, et permettent ainsi de décrire une relation de forme complexe avec très peu de segments – et donc de degrés de liberté –, tout en réduisant l'influence du choix arbitraire de la position des nœuds.

La seconde méthode alternative consiste à introduire dans le modèle la valeur de la variable explicative mise au carré et éventuellement au cube. Cette approche a été critiquée pour sa non localité : elle s'adapte mal à des changements abrupts de l'effet, et une modification dans une partie de l'échelle des valeurs peut avoir d'importantes répercussions à son autre extrémité. Au contraire, grâce au découpage en segments, les splines représentent bien les transformations locales de la forme de l'effet.

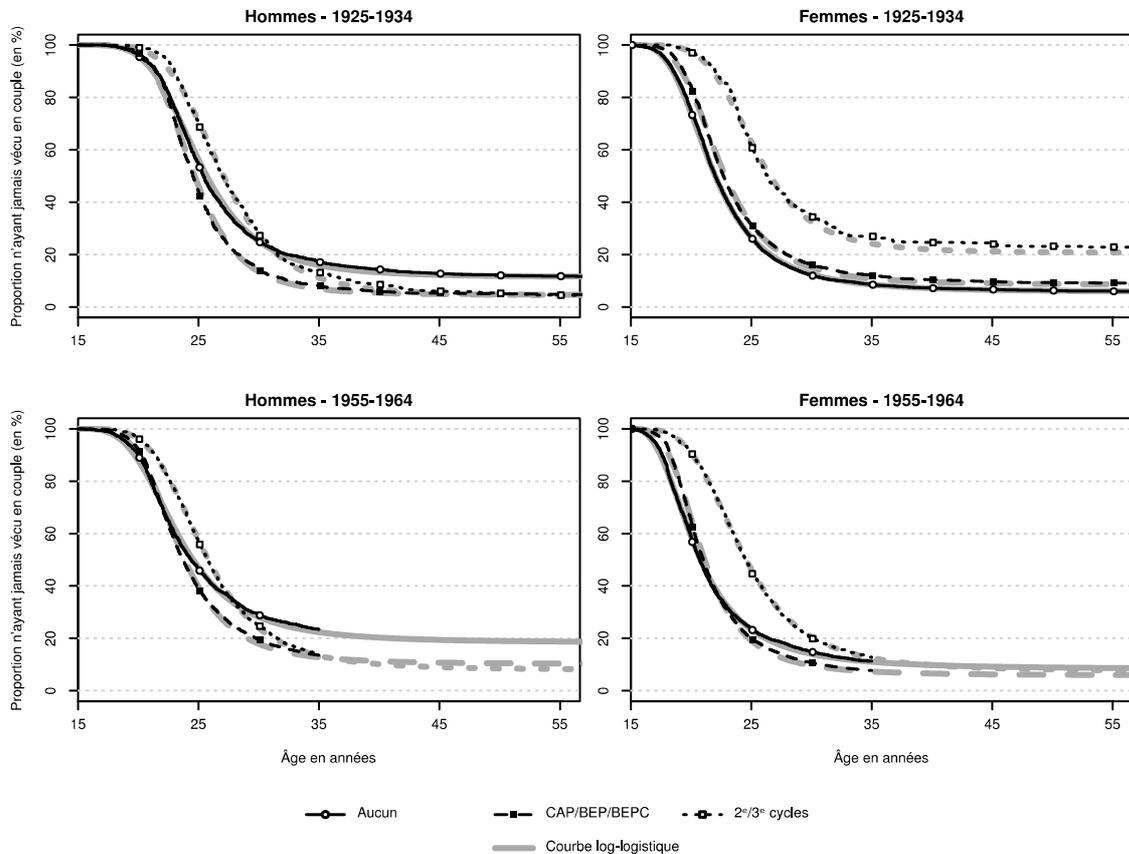
En pratique, les splines cubiques restreintes sont introduites dans le modèle sous forme de nouvelles variables explicatives (dites bases) calculées à partir de la variable originale. Chaque segment donne lieu à une variable numérique supplémentaire, et le coefficient associé mesure la déviation par rapport à la relation modélisée par les précédents. Présentées plus bas dans les tableaux dans leur forme brute, les splines peuvent être reformulées sous une forme plus longue comprenant un coefficient par segment, plus deux coefficients décrivant les effets linéaires applicables chacun au-delà des nœuds extrêmes de chaque côté de l'échelle. La complexité des informations à transmettre appelle une représentation graphique, que nous privilégions systématiquement.

3.2 L'adéquation du modèle aux données

Nous commençons par valider l'application de la distribution log-logistique avec immunité au cas de la première mise en couple. Afin d'illustrer à la fois l'adéquation de la distribution aux données et la forme qu'elle prend pour différents niveaux de diplôme, nous présentons ici les courbes de séjour de deux cohortes pour trois niveaux de diplôme. L'interprétation des résultats et la présentation de l'évolution temporelle seront réalisées dans la section suivante, à travers l'étude de la valeur des paramètres au fil des cohortes.

L'adéquation de la distribution aux données s'avère très bonne pour tous les groupes de diplôme des deux sexes et des deux générations (Figure 5.3). En particulier, la distribution décrit très bien l'évolution de la proportion de personnes ayant vécu en couple au-delà de 55 ans pour la première génération, ce qui renforce la crédibilité de l'extrapolation réalisée concernant la seconde génération (courbes grises). Si l'on observe un léger décalage entre 35 et 55 ans parmi les femmes de la première génération diplômées des deuxième et troisième cycle du supérieur, la courbe de séjour log-logistique rejoint finalement la courbe empirique au-delà, et reflète donc correctement le taux de célibataires définitives de ce groupe ; rappelons par ailleurs que les effectifs de ce groupe dans l'échantillon sont assez réduits pour cette génération (400 personnes au total, soit environ 100 encore célibataires à 35 ans), ce qui peut expliquer l'existence de quelques déviations aléatoires. La qualité de l'ajustement de la courbe log-logistique n'est pas étonnante : elle a déjà été appliquée à cet objet par plusieurs travaux (Keeley, 1979 ; Brüderl et Diekmann, 1995, 1997) ; elle est en outre très proche de la distribution Coale-McNeil classiquement utilisée pour l'étude du mariage (Coale et McNeil, 1972 ; Kaneko, 2003).

Figure 5.3 : Distribution des âges de mise en couple (courbe de séjour) pour trois groupes de diplômés au sein de deux générations



Les courbes noires correspondent à l'estimateur non paramétrique de Kaplan-Meier, les courbes grises à la prédiction d'un modèle log-logistique autorisant l'âge médian de mise en couple, la forme de la distribution et la proportion finale de mises en couple à varier selon le diplôme.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France

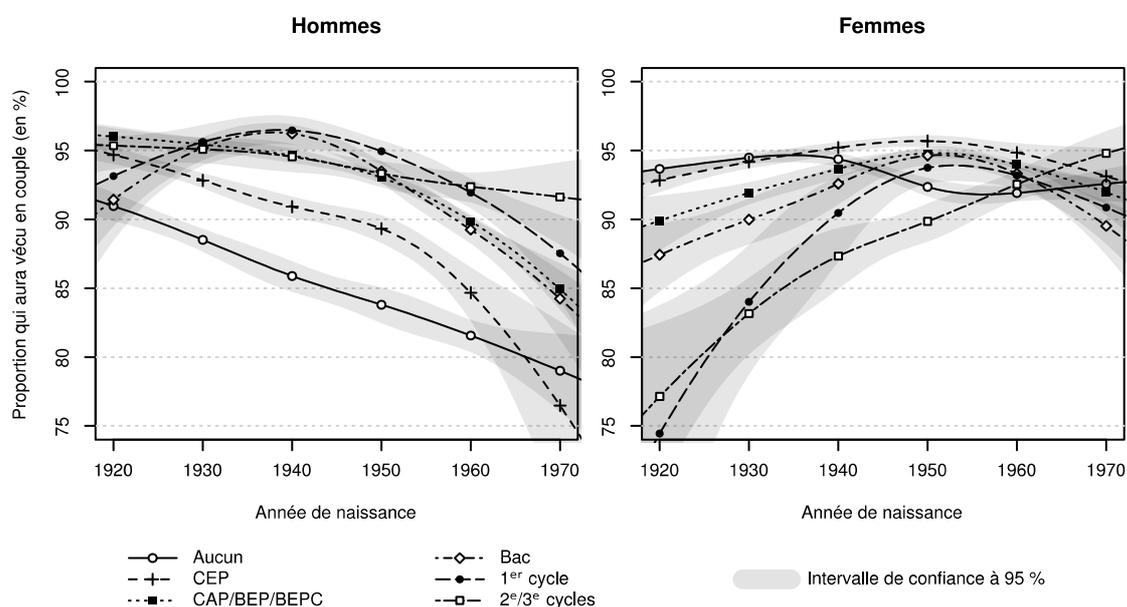
4 L'évolution et les mécanismes du célibat définitif

4.1 La résorption du surcroît de célibat des diplômées

Les résultats du modèle log-logistique avec immunité confirment l'existence d'un modèle fortement genré de mise en couple parmi les générations d'avant-guerre, déjà observé au chapitre précédent pour l'année 1975. Chez les hommes, seuls les non diplômés, et moins nettement les titulaires du certificat d'études primaires (CEP), se distinguent par des chances moins élevées de vivre en couple, inférieures respectivement de 5 et 10 points

de pourcentage pour la cohorte 1940 (Figure 5.4) ; aucune différence ne s'observe entre les autres niveaux de diplôme. Chez les femmes au contraire, on observe une échelle très nette, les chances de vivre en couple diminuant régulièrement avec l'augmentation du niveau de diplôme (passant ainsi de 95 % à 75 %). Ce tableau correspond parfaitement à une situation dans laquelle l'hypergamie prévaut.

Figure 5.4 : Proportion d'individus qui auront vécu au moins une fois en couple selon l'année de naissance et le diplôme



Note : l'apparente convergence dans les années 1920 entre la courbe correspondant aux hommes les moins diplômés et les autres courbes semble devoir être attribuée à un effet de sélection découlant des différences de mortalité selon le milieu social au-delà de 70 ans (Blanpain, 2011).

Un modèle log-logistique autorisant l'âge médian de mise en couple, la forme de la distribution et la proportion finale de mises en couple à varier selon le diplôme est utilisé.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France

Mais ici encore, ce modèle de mise en couple n'a plus cours parmi les générations les plus récentes, et l'évolution au fil des cohortes des différences entre groupes de diplôme est très nette. Chez les hommes, le célibat définitif se développe parmi les moins diplômés dès la cohorte 1920, puis dans tous les groupes à partir de la cohorte 1940, en affectant cependant moins fortement les diplômés les plus élevés. Chez les femmes, les différences entre groupes

de diplôme se résorbent de manière spectaculaire et disparaissent complètement. Les chances de vie en couple des plus diplômées rejoignent celles des moins diplômées dès la cohorte 1950, à l'exception des diplômées des deuxième et troisième cycle du supérieur, pour lesquelles le rattrapage est un peu plus tardif.

Ainsi, nous observons comme Maria Winkler-Dworak et Laurent Toulemon (2007, p. 296) une diminution des chances de mise en couple des non diplômées par rapport aux diplômées ; mais c'est l'augmentation des chances de mise en couple de ces dernières qui nous semble constituer la principale évolution chez les femmes¹³. Par ailleurs, nous confirmons leur observation selon laquelle les différences entre hommes et femmes se sont résorbées pour tous les groupes de diplôme, à l'exception des hommes non diplômés, chez qui les chances de mise en couple restent plus faibles.

Ces résultats concernant le célibat définitif contrastent en partie avec ceux portant sur la mesure du célibat à un instant donné obtenus au chapitre précédent (Figure 271, p. 4.4) : dans ce dernier cas, la convergence était réalisée par la diminution du taux de personnes en couple parmi les moins diplômées. Dans les deux cas, les différences entre groupes évoluent de la même manière, et seule la dynamique d'ensemble diffère : diminution de la proportion de personnes qui ne vivront jamais en couple d'un côté, augmentation du taux de personnes qui vivent en couple à un instant donné de l'autre. En outre, le célibat définitif des femmes non diplômées, s'il a légèrement augmenté, est encore nettement inférieur à celui des hommes du même groupe ; au contraire, le célibat à un instant donné est désormais au même niveau chez les deux sexes. L'augmentation de la fréquence des séparations et des remises en couple a rendu possible ce décalage entre les deux types de célibat.

Les évolutions du célibat définitif peuvent être directement mises en relation avec celles de l'hypergamie commentées dans la section précédente. En effet, c'est avec les cohortes nées aux alentours de 1945 que s'enclenchait la baisse de l'hypergamie féminine, en termes absolus comme relatifs. Ce sont ici aussi ces cohortes qui marquent chez les femmes la disparition presque complète des différences de célibat définitif liées au diplôme, et qui initient chez les hommes le mouvement d'augmentation du célibat définitif pour tous les groupes de diplôme. Ainsi se trouve à nouveau confirmée la correspondance fondamentale

13 Cette évolution a aussi été observée par Isabelle Robert-Bobée et Magali Mazuy (2005, p. 185) à l'âge de 45 ans pour les cohortes nées avant 1955.

entre hypergamie féminine et célibat des diplômées avancée au chapitre précédent : le désavantage des diplômées sur le marché conjugal a disparu en même temps que le déséquilibre créé par la préférence pour l'hypergamie féminine.

On peut penser que la massification scolaire a joué un grand rôle dans cette évolution. D'une part, le célibat des diplômées du supérieur au sein des cohortes d'avant-guerre ne condamnait pas une proportion importante des hommes au célibat, du fait des faibles effectifs de ce groupe chez les femmes. Au contraire, au sein des générations les plus récentes, la persistance d'un niveau si élevé de célibat chez les diplômées n'était pas tenable, tant elle aurait profondément déséquilibré le marché conjugal, forçant les hommes diplômés à prendre pour conjointes des femmes issues d'un milieu social très éloigné du leur, et condamnant l'ensemble des non diplômés au célibat. D'autre part, les diplômées du supérieur nées avant-guerre ne peuvent être directement comparées à leurs cadettes nées après 1950 : les premières constituaient une population fortement sélectionnée qui avait effectué le choix relativement transgressif de faire passer les études et la carrière avant la formation d'une famille. Comme on le voit, ce conflit a perdu en importance avec la banalisation des études féminines.

L'augmentation du célibat des hommes peu diplômés, opposé à une relative stabilité chez les femmes de même niveau de diplôme, vient nuancer le constat d'une convergence entre les deux sexes. Alors que les chances de vie en couple ne varient plus selon le diplôme parmi les femmes des cohortes récentes, on note chez les hommes l'apparition et le développement d'une hiérarchie entre diplômés : les chances de vivre en couple s'élèvent désormais avec le niveau d'éducation. Cette échelle est l'exact inverse de celle qui prévalait chez les femmes nées avant guerre. Loin d'indiquer une inversion des positions des hommes et des femmes, ainsi que nous avons pu l'observer plus haut à propos de l'hypergamie, elle révèle plutôt la persistance d'un modèle genré dans lequel le statut social de l'homme importe plus que celui de sa conjointe – les femmes non diplômées ne connaissant pas les mêmes difficultés. En effet, la massification scolaire a donné de plus en plus d'importance au diplôme sur le marché du travail, concentrant sur les moins diplômés – et en particulier les non diplômés, moins nombreux qu'auparavant – les risques d'une insertion professionnelle difficile et d'une précarité durable (Galland, 2000 ; Gasquet, 2003). Ces difficultés affectent particulièrement la mise en couple des hommes en ralentissant leur départ du foyer

parental, quand les femmes se replient au contraire sur leur vie conjugale ou familiale (Galland, 1995 ; Jaspard, Brown et Battagliola, 1995 ; Dufour-Kippelen et Dormont, 2000 ; Solaz, 2000). La situation des hommes peu diplômés évoque à ce titre celle des petits agriculteurs au sein des générations précédentes ; on est tenté de faire le rapprochement avec l'observation de Guenhaël Jegouzo (1972, p. 13) selon laquelle le célibat masculin « *est un des signes les plus manifestes de la dégradation de la condition paysanne* »¹⁴.

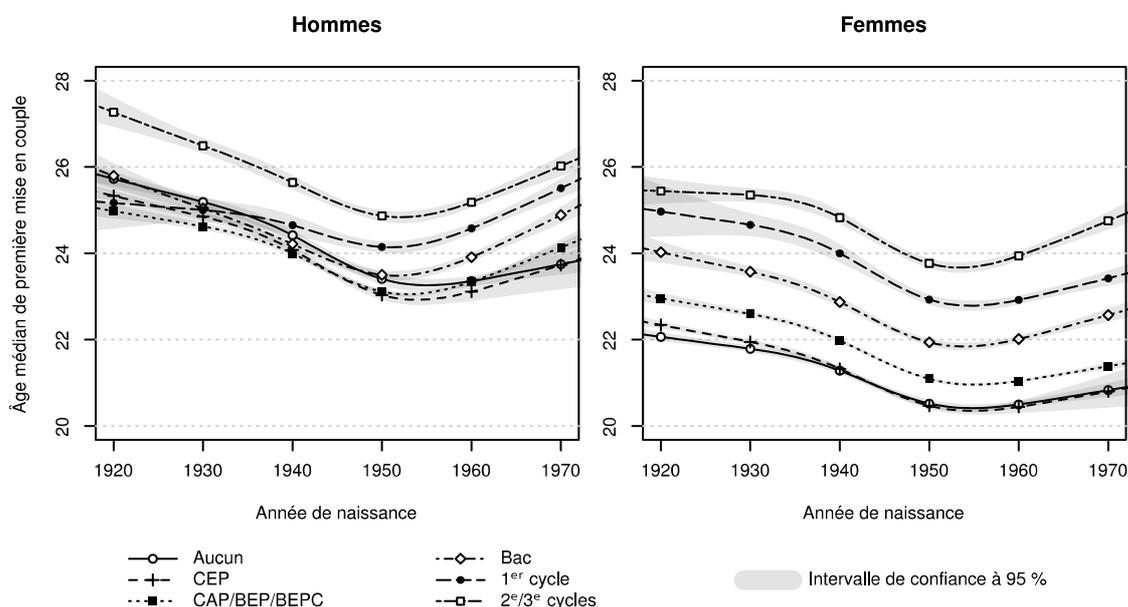
Néanmoins, le contraste avec la réduction des écarts liés au diplôme du point de vue du célibat à un instant donné observée au chapitre précédent interroge. La différence assez importante entre les taux de célibat définitifs masculins et féminins est elle aussi assez étonnante, impliquant des remises en couple nombreuses parmi les hommes qui ne restent pas définitivement célibataires. Rappelons que les estimations du célibat définitif portant sur les cohortes postérieures à 1960 doivent être considérées avec prudence, les interrogés ayant moins de 30 ans à la date de l'enquête. Ces résultats ne pourront être confirmés qu'à l'aide de données plus récentes – potentiellement fournies par de futures enquêtes Famille qui intégreraient à nouveau une question portant sur la première mise en couple.

4.2 Le renforcement du lien entre diplôme et âge de mise en couple chez les hommes

Le mécanisme envisagé par F. de Singly expliquait l'existence d'un fort célibat des diplômées par le décalage entre le calendrier des études et celui de la mise en couple. On peut penser que la résorption du célibat de ce groupe est liée à la disparition de ce désajustement du marché conjugal. En particulier, une élévation de l'âge médian de mise en couple de l'ensemble des femmes, le rapprochant de celui des hommes, laisserait aux plus diplômées le temps nécessaire pour terminer leurs études avant de se mettre en couple, sans souffrir de la concurrence des moins diplômées. Mais cette convergence n'a pas eu lieu : si l'on observe un rapprochement partiel entre hommes et femmes des *variations selon le diplôme* dans l'âge de première mise en couple, *l'écart entre les sexes* n'a pas diminué.

14 De fait, le groupe des non diplômés et celui des (petits) agriculteurs se recoupaient largement parmi les générations d'avant-guerre (Jegouzo et Dauce, 1969, p. 52).

Figure 5.5 : Âge médian de première mise en couple selon l'année de naissance et le diplôme



L'âge médian est celui des seuls individus qui se mettront au moins une fois en couple. Un modèle log-logistique autorisant l'âge médian de mise en couple, la forme de la distribution et la proportion finale de mises en couple à varier selon le diplôme est utilisé.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France

Ainsi, les évolutions les plus notables de l'âge médian de mise en couple sont communes aux hommes et aux femmes et à tous les niveaux de diplôme : elles n'ont pas affecté l'écart entre les sexes, et peu l'écart entre les différents groupes (Figure 5.5). L'âge de première mise en couple connaît une première phase de baisse jusqu'aux cohortes nées autour de 1955, pour augmenter depuis (Prioux, 2003 pour la France ; Blossfeld, 1995, p. 17 pour d'autres pays occidentaux)¹⁵. Pour chaque niveau de diplôme, l'âge médian de première mise en couple reste pour la cohorte 1970 inférieur à ce qu'il était pour la cohorte 1930 ; néanmoins, du fait de l'augmentation générale du niveau d'éducation, les âges médians observés dans ces deux cohortes sont très proches (Prioux, 2003).

Chez les femmes, ces évolutions d'ensemble ont affecté de manière parallèle tous les groupes : l'âge médian de première mise en couple suit parfaitement le niveau de diplôme

15 Le lien entre conjoncture économique favorable et mise en couple et émancipation précoces se retrouve dans de nombreuses configurations historiques (Reher, 1998, p. 208).

pour toutes les cohortes. Chez les hommes, en revanche, des divergences peuvent être relevées. En effet, on n'observait guère d'écarts entre les groupes de diplôme au sein des générations d'avant-guerre, à l'exception des diplômés des deuxième et troisième cycles du supérieur, qui se mettaient en couple plus tardivement. Mais une échelle s'est progressivement affirmée, en particulier du fait que les hommes non diplômés se sont mis en couple de plus en plus tôt, pour devenir finalement plus précoces que tous les autres groupes¹⁶. Ainsi, pour les cohortes les plus récentes, on retrouve la même échelle chez les hommes et chez les femmes, même si les écarts restent bien plus marqués chez ces dernières (Galland, 1995, p. 47).

Au total, le décalage des calendriers masculin et féminin persiste donc : il ne semble plus possible de considérer qu'il ait été à l'origine du fort célibat des diplômées, ce dernier phénomène ayant disparu. Néanmoins, les profondes transformations du couple qui ont pris place au fil des cohortes impliquent que ce décalage a pu cesser d'être une difficulté pour les femmes les plus diplômées. Le développement des cohabitations hors mariage et de la contraception (Leridon *et al.*, 1987 ; Guibert-Lantoine et Leridon, 1998) permet plus facilement aux cohortes récentes de concilier vie conjugale et études. De fait, comme nous le verrons au chapitre suivant, alors que pour les cohortes d'avant-guerre la mise en couple en cours de scolarité était très rare, cette pratique s'est progressivement répandue chez les diplômés du supérieur. Par ailleurs, l'augmentation de la fréquence des séparations et des divorces a probablement atténué la division entre un « marché primaire » majoritaire et un « marché secondaire » très atypique, facilitant ainsi une première mise en couple relativement tardive avec des individus ayant déjà connu une expérience conjugale.

Le délai requis pour décrocher un diplôme élevé ne pourrait donc constituer une entrave à la mise en couple des femmes que dans le cadre d'un modèle relativement rigide dans lequel l'homme et la femme doivent avoir achevé leurs études pour être en mesure d'assurer les rôles genrés qui leur sont respectivement assignés. Cette hypothèse est-elle confirmée par l'étude de la distribution des âges de mise en couple ?

16 Et ce, malgré un âge de fin d'études médian qui augmente plus nettement que pour les autres groupes, passant de 14 ans pour la cohorte 1920 à 17 ans pour la cohorte 1970.

4.3 L'allongement de la période de mise en couple des femmes

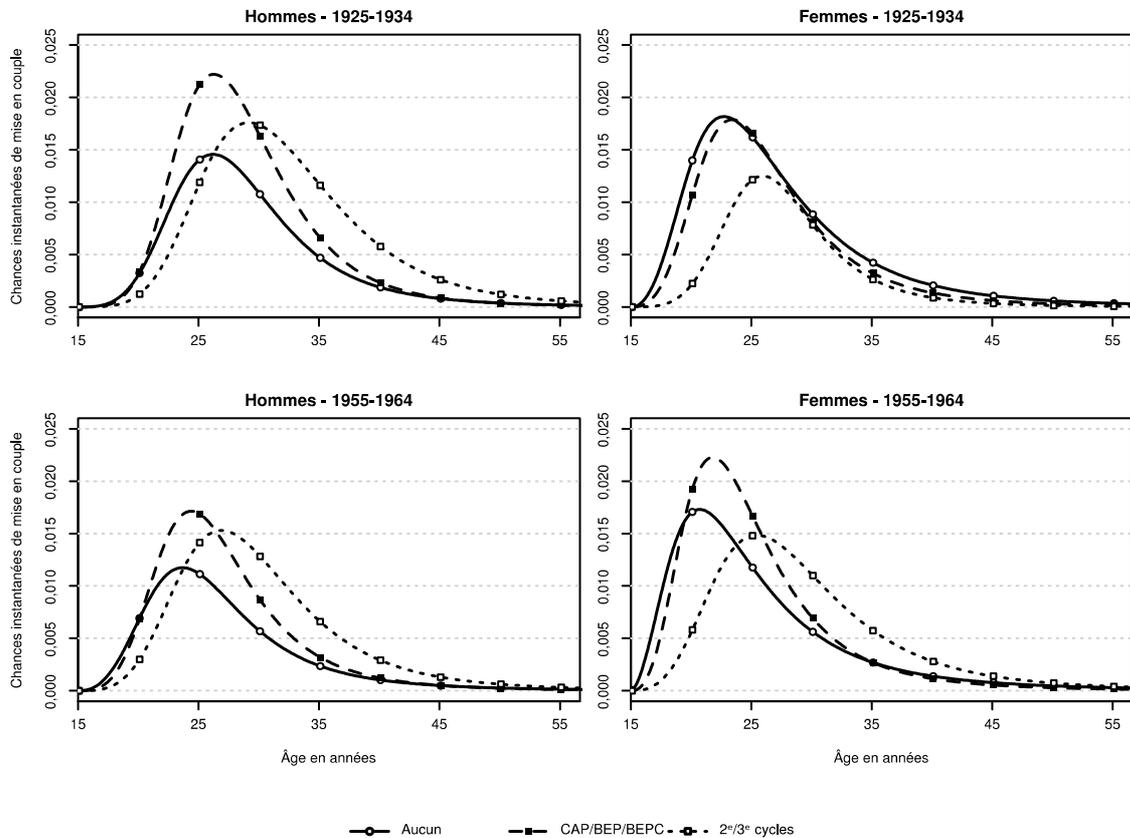
Nous avons jusqu'ici passé en revue les deux premiers paramètres du modèle log-logistique (proportion de célibataires définitifs et âge médian de mise en couple), mais nous n'avons pas observé le profil complet des âges de mise en couple. Celle-ci dépend aussi du troisième paramètre du modèle (forme de la distribution), que nous avons laissé de côté jusqu'ici étant donné la difficulté à l'interpréter de manière isolée.

Nous présentons dans la figure 5.6, pour deux cohortes et trois groupes de diplôme, les courbes de chances instantanées de mise en couple, qui mesurent les chances de se mettre en couple à un instant donné pour les individus n'ayant encore jamais vécu en couple à cet âge. Les trois paramètres du modèle trouvent leur traduction sur ces courbes : l'âge médian de mise en couple correspond au décalage du pic de la courbe sur l'axe des abscisses ; la proportion d'individus qui vivront en couple correspond à l'aire sous la courbe¹⁷ ; la forme de la distribution correspond à la plus ou moins forte concentration des mises en couple et à leur plus ou moins grande symétrie autour du pic.

Si le célibat des femmes les plus diplômées nées avant guerre s'explique bien par un décalage entre calendrier des études et calendrier de mise en couple, nous devrions observer une augmentation (trop) tardive des chances de mise en couple des femmes de ce groupe. Est-ce le cas ? Pour toutes les cohortes, et chez les deux sexes, le début de la phase de croissance des chances de mise en couple a lieu à peu près au même âge pour les non diplômés et les titulaires d'un CAP, d'un BEPC ou d'un BEPC, mais il a lieu légèrement plus tard chez les diplômés des deuxième et troisième cycle du supérieur. À l'appui de l'explication que nous venons de proposer, il est particulièrement tardif chez les femmes diplômées du supérieur nées entre 1925 et 1934, ce que l'on peut interpréter comme le signe de l'incompatibilité entre études et mise en couple, qui a diminué pour les cohortes les plus récentes.

17 Formellement, en notant l'aire sous la courbe a , ce taux s'écrit $k = 1 - \exp(-a)$. L'aire est elle-même égale à la limite en $+\infty$ de la fonction de chances cumulée.

Figure 5.6 : Chances instantanées de mise en couple pour trois groupes de diplômes au sein de deux générations



Les courbes correspondent à la prédiction d'un modèle log-logistique autorisant l'âge médian de mise en couple, la forme de la distribution et la proportion finale de mises en couple à varier selon le diplôme.

Note : Ces courbes ne doivent pas être confondues avec les fonctions de densité de probabilité. En particulier, l'âge médian de mise en couple ne correspond pas, comme on peut être tenté de le croire, à celui qui divise l'aire sous une courbe en deux zones de taille égale. En effet, puisque les individus sortent de la population « à risque » une fois qu'ils se sont mis en couple, les effectifs que représente une aire donnée diminuent progressivement à mesure que l'âge augmente.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France

Mais ce sont les différences concernant la phase de diminution des chances de mise en couple qui sont les plus frappantes. Alors que chez les hommes de toutes les cohortes, et chez les femmes nées après guerre, cette phase est de plus en plus tardive à mesure que le diplôme augmente, chez les femmes nées entre 1925 et 1934, la décline des chances de mise en couple a lieu simultanément chez tous les groupes de diplôme. Passé 25 ans, soit

traditionnellement l'âge de « coiffer Sainte Catherine », les chances de mise en couple diminuaient rapidement : tout se passe comme si pour les femmes de cette génération, la mise en couple n'était plus envisageable au-delà de 35 ans.

On peut noter un phénomène symétrique chez les hommes les moins diplômés, en particulier ceux nés entre 1955 et 1964 – parmi lesquels nous avons vu que le célibat définitif était le plus fort –, et, quoique moins nettement, chez les femmes du même groupe nées entre 1955 et 1964 : leurs chances de mise en couple commencent à augmenter en même temps que celles des diplômés du CAP, du BEP ou du BEPC, mais elles entament leur déclin beaucoup plus tôt. On peut faire l'hypothèse que des difficultés d'insertion professionnelle sont au-delà d'un certain âge une entrave à l'entrée en cohabitation.

Ainsi, il apparaît que les groupes les plus exposés au célibat définitif ne souffrent pas d'une difficulté uniforme à se mettre en couple, mais d'une quasi-impossibilité à le faire à certains âges. Chez les femmes les plus diplômées nées avant guerre, la difficulté à se mettre en couple aussi bien tôt que tard entraînait une réduction importante des chances de vie en couple. Chez les moins diplômés, hommes, mais aussi à un moindre degré femmes, des cohortes récentes, c'est la difficulté à se mettre en couple tardivement qui pose problème.

La résorption du surcroît de célibat des femmes diplômées est donc due en partie à une diminution de l'incompatibilité entre études et vie en couple (phénomène sur lequel nous reviendrons au chapitre suivant), mais surtout à la disparition d'une barrière concernant l'âge maximum de mise en couple. L'hypothèse de F. de Singly (1987, chap. 8) concernant le décalage entre calendrier des études et de la mise en couple est donc validée. L'existence de cette barrière nous semble à son tour s'expliquer par la prévalence de l'hypergamie parmi les générations d'avant-guerre : c'est parce que les hommes ne se mettaient pas en couple en priorité avec les femmes les plus diplômées, mais laissaient leur chance à tous les âges aux moins diplômées, que les premières, arrivées trop tard sur le marché conjugal, se trouvaient désavantagées. Ce que nous enseigne en effet la comparaison des évolutions avec l'âge des chances de mise en couple par diplôme, c'est que, même après 30 ans, les femmes les plus diplômées voyaient leurs chances demeurer encore inférieures à celles des moins diplômées – ce qui n'est plus le cas parmi les générations récentes. La position de la « limite de la Sainte Catherine » à 25 ans est liée au fait que, l'écart d'âge entre conjoints étant parmi les générations d'avant-guerre d'environ 3 ans (Vanderschelden, 2006b ; Mignot, 2010, p. 291),

les chances de mise en couple des femmes étaient très faibles au-delà de l'âge auquel les hommes les plus diplômés – à la mise en couple la plus tardive – s'étaient typiquement mis en couple.

Mais la préférence pour l'hypergamie féminine n'explique sans doute pas à elle seule l'intégralité du surcroît de célibat des diplômées. En effet, la tendance à former des couples dans lesquels la femme est plus âgée que l'homme implique qu'à moins d'attendre longtemps après la fin de leurs études, les hommes ne peuvent prendre pour conjointe une femme préparant un diplôme supérieur ou égal au leur dans un contexte d'incompatibilité entre études et mise en ménage : même en l'absence d'une préférence pour l'hypergamie féminine, les plus diplômées se trouvaient en position défavorable sur le marché conjugal. Le développement de la cohabitation hors mariage a atténué cette incompatibilité, rendant les mises en couple en cours d'études plus fréquentes (*cf.* chapitre suivant), tout en contribuant à réduire l'écart d'âge entre conjoints (Vanderschelden, 2006b) : le conflit entre calendriers a nettement perdu en intensité pour les femmes, participant à la diminution du célibat des diplômées.

5 Conclusion

Confirmant les résultats du chapitre précédent, nous avons mis en évidence une baisse, puis une inversion de l'hypergamie féminine de diplôme parmi les premiers couples cohabitants entre les générations nées avant la Seconde Guerre mondiale et celles nées dans les années 1970. Ce mouvement a affecté les trois dimensions de l'hypergamie : hypergamie absolue, hypergamie relative, et structure genrée du célibat selon le diplôme. Premièrement, suite à l'élévation du niveau d'éducation des femmes, qui dépasse désormais celui des hommes, c'est plus souvent la femme que l'homme qui est la plus diplômée au sein des premiers couples depuis les cohortes nées à la fin des années 1950. Deuxièmement, même en contrôlant l'évolution de la structure par diplôme de la population, on observe une inversion similaire de l'hypergamie relative. Ce résultat indique qu'au-delà des contraintes posées par la disponibilité de conjoints d'un niveau de diplôme donné, les individus des cohortes récentes « choisissent » plus souvent de former un couple dans lequel la femme est la plus diplômée que l'inverse, par rapport à ce qui aurait lieu si la mise en couple avait lieu au hasard. Troisièmement, les taux de célibat définitif des femmes, qui croissaient nettement

avec le niveau de diplôme parmi les générations d'avant-guerre, se sont rapprochés au point qu'on n'observe plus aucune différence parmi les générations nées dans les années 1960. Ces trois résultats confirment ceux du chapitre précédent, tout en précisant la chronologie historique des évolutions : c'est avec les premières cohortes du *baby-boom*, nées au milieu des années 1940, que s'enclenche le mouvement d'inversion de l'hypergamie féminine, qui était jusqu'alors relativement stable. Cette temporalité rattache l'évolution de l'hypergamie aux grandes transformations qui ont affecté la famille à partir de ces générations-charnières (Bonvalet, Clément et Ogg, 2011).

Ces résultats confirment aussi le lien entre célibat et hypergamie féminine, dont l'étude de l'évolution avec l'âge des chances de mise en couple par niveau de diplôme précise le mécanisme. La barrière symbolique des 25 ans, correspondant à l'âge de « coiffer Sainte Catherine », marquait pour les femmes des générations nées avant guerre une diminution rapide des chances de mise en couple. Les plus diplômées, qui avaient plus fréquemment que les autres repoussé cet événement, se trouvaient en difficulté sur le marché conjugal, les hommes leur ayant préféré des femmes moins diplômées et plus jeunes, disposées à se mettre en couple plus tôt. La fin de l'hypergamie féminine de diplôme, le développement des cohabitations facilitant la mise en couple en cours d'études et la réduction de l'écart d'âge en faveur de l'homme ont eu raison de cette situation dès les cohortes nées dans les années 1950.

Cependant, des différences entre hommes et femmes persistent. D'une part, la première mise en couple reste fortement genrée, même au sein des cohortes récentes. Nous en avons vu deux signes : les âges médians de première mise en couple des hommes et femmes restent très différents ; le célibat définitif demeure supérieur parmi les hommes à ce qu'il est parmi les femmes ne possédant aucun diplôme. Ce dernier résultat est en contradiction avec la convergence observée en termes de célibat à un instant donné. Il pourrait indiquer une divergence réelle, les femmes non diplômées ayant tendance à se mettre fréquemment en couple de manière à quitter tôt le foyer parental, mais à se séparer plus souvent ou à moins fréquemment se remettre en couple suite à une séparation. Elles sont par exemple plus fréquemment à la tête d'une famille monoparentale que les autres femmes (Prioux, Mazuy et Barbieri, 2011, p. 445), ce qui peut s'expliquer par une plus grande précocité de la conception, la présence d'enfants diminuant les probabilités de remise en couple (Cassan,

Mazuy et Clanché, 2005). Les effets opposés de la parentalité sur la remise en couple des hommes et des femmes expliqueraient ainsi la divergence entre célibat définitif et célibat à un instant donné selon le sexe. La convergence des taux de célibat à un instant donné des hommes et femmes non diplômés masquerait alors des différences de situation nettement genrées au sein de ce groupe : insertion professionnelle laborieuse pour les hommes, difficultés à concilier parentalité et vie conjugale pour les femmes. Cette hypothèse, qui reste à creuser, apporterait une nuance au constat d'un rapprochement des facteurs de vie en couple des hommes et des femmes, sans pour autant le remettre en question concernant l'ensemble des groupes de diplôme.

CHAPITRE 6

LES DÉTERMINANTS BIOGRAPHIQUES DE L'HOMOGAMIE ET DE L'HYPERGAMIE DE DIPLÔME AU SEIN DES PREMIERS COUPLES : L'ÂGE ET LA SCOLARISATION (COHORTES 1920-1970)

Par contraste avec le corpus important de travaux mesurant la structure, les évolutions et les conséquences de l'homogamie sociale dans de nombreux pays que nous avons présenté en première partie, on sait encore relativement peu de choses des déterminants du choix du conjoint. Parmi différentes approches possibles des mécanismes influençant ce choix, nous adoptons dans ce chapitre celle de l'analyse biographique. Après quelques travaux pionniers (Lichter, 1990 ; Mare, 1991 ; Smeenk, 1998 ; Blossfeld et Timm, 2003a), cette veine de recherche semble s'être quelque peu tarie. Nous voudrions ici reprendre cette inspiration pour l'appliquer au cas de la France, à partir de données plus riches et de modèles que nous jugeons plus appropriés que ceux utilisés par la plupart des travaux précédents, et tenter de montrer ainsi que cette approche peut encore se révéler fructueuse pour l'explication du choix du conjoint.

L'objet de ce chapitre est en particulier d'utiliser l'analyse biographique afin de mieux comprendre les évolutions de l'homogamie au cours du temps observées dans les chapitres précédents. Nous retiendrons donc ici un double questionnement. D'un côté, nous nous demanderons comment expliquer l'affaiblissement de l'homogamie de diplôme observé en France depuis 40 ans observé aux chapitres 1 et 2, alors qu'on aurait pu s'attendre à ce que la massification scolaire qui a pris place sur cette période entraîne au contraire une augmentation de celle-ci – selon le mécanisme prévu par Hans-Peter Blossfeld et Andreas Timm (2003a). De l'autre, nous chercherons à observer si l'inversion en faveur des femmes de l'hypergamie de diplôme mise en évidence aux chapitres 4 et 5 s'est accompagnée d'une transformation des différences genrées dans le processus de première mise en couple et de

choix du conjoint, et ce que ces différences nous disent quant à la réalité de la norme d'hypergamie féminine. Plus largement, ces deux interrogations constituent une entrée pour la compréhension des processus de mise en couple et du fonctionnement du marché conjugal.

Parmi les éléments de contexte de la mise en couple qui semblent pouvoir affecter le choix du (premier) conjoint en termes de diplôme, deux aspects biographiques paraissent fondamentaux : l'âge de mise en couple, et la situation par rapport aux études. Ces deux dimensions ont en commun d'affecter fortement les opportunités de rencontre des individus. Nous faisons ici l'hypothèse – largement confirmée dans ce qui suit – que la structure du marché conjugal dans lequel évoluent les individus est un déterminant essentiel du choix du conjoint (Blau et Schwartz, 1984 ; Blau, 1994), déterminant plus fort que d'éventuelles stratégies individuelles ou qu'une possible norme d'hypergamie féminine.

Le premier élément de contexte que nous étudierons est l'âge de l'individu. Indissociablement biologique et social, l'âge est un indicateur indirect de la situation biographique et du statut social. Chez les actifs, il est le signe d'une carrière professionnelle plus avancée qui peut être valorisée sur le marché conjugal. Les effets potentiellement différents de ce facteur sur le choix du conjoint des hommes et des femmes, et l'évolution de cette différence au fil des cohortes, retiendront particulièrement notre attention. Mais l'âge ne définit pas seulement les caractéristiques individuelles : il affecte aussi et surtout les opportunités de rencontre sur le marché conjugal. En effet, à mesure que l'âge d'un individu augmente, la proportion de célibataires au sein de sa cohorte de naissance diminue, et la répartition en termes de diplôme des membres de cette cohorte n'ayant jamais vécu en couple se modifie. Après avoir étudié cette contrainte structurelle et ses déterminants, nous tenterons de mesurer son influence sur le choix du conjoint.

La seconde dimension qui apparaît essentielle dans le contexte de la mise en couple est la situation de l'individu par rapport aux études et au marché du travail. En effet, le contexte scolaire met en relation des individus d'âge proche, et sélectionnés par leur progression dans le système scolaire même s'ils ne poursuivront pas tous leurs études jusqu'au même niveau : il constitue un marché conjugal local particulièrement efficace. Au contraire, on peut penser que le milieu professionnel met plus souvent en contact des individus occupant des postes différents dans la division sociale du travail, et ayant suivi des formations plus

hétérogènes ; mais surtout que la plus grande disparité en termes d'âge y limite les possibilités de rencontres amoureuses, et force les individus à trouver des conjoints dans d'autres lieux plus mixtes. Il semble donc légitime de se demander si les individus qui se mettent en couple en cours d'études, ou rapidement après les avoir achevées, présentent un profil particulier du point de vue du choix du conjoint, et si ce profil varie selon le sexe et les générations. De ce point de vue, l'augmentation du niveau d'éducation, en particulier celui des femmes, constitue l'évolution majeure susceptible d'avoir affecté les processus de choix du conjoint.

Ainsi, il s'agira dans ce chapitre de déterminer si, à diplôme final contrôlé, le choix du conjoint en termes de diplôme évolue nettement selon l'âge et les circonstances de la mise en couple. En particulier, nous nous demanderons si la mise en couple pendant ou à proximité des études est un facteur de renforcement de l'endogamie ou de l'hypergamie de diplôme. Enfin, nous tenterons de déterminer si les effets des circonstances de la mise en couple sont différents chez les hommes et les femmes, et si l'on observe ou non une convergence au fil des cohortes.

Nous commençons par passer en revue les mécanismes théoriques développés dans la littérature, ainsi que les quelques résultats empiriques déjà établis. Puis nous expliquerons comment les enquêtes Étude de l'histoire familiale (Insee-Ined, 1999) et Biographies et entourage (Ined, 2001), déjà présentées au chapitre précédent, peuvent être utilisées pour la présente étude. L'analyse sera ensuite conduite en deux temps.

Nous chercherons dans un premier temps à évaluer l'importance numérique et la signification des mises en couple pendant ou peu après les études. Nous montrerons que, du fait notamment de la massification scolaire, la fréquence des mises en couple en cours d'études a nettement augmenté chez les deux sexes, et que la durée écoulée entre fin des études et mise en couple s'est parallèlement réduite. Nous verrons que les couples qui se sont formés pendant les études des deux conjoints sont plus souvent endogames que les autres, mais que cette différence ne s'étend pas aux couples formés pendant la scolarisation d'un seul des deux conjoints.

Dans un second temps, nous développerons un modèle à risques concurrents avec variables dépendantes du temps, et tenterons de répondre aux questions posées ci-dessus. Nos résultats tendent à confirmer les travaux précédents concernant les effets de la

scolarisation ; ils mettent aussi en évidence un effet très fort, et peu étudié jusqu'ici, de l'âge de mise en couple, de même que des évolutions assez nettes au fil des cohortes. D'une part, la scolarisation augmente les chances d'endogamie et d'hypergamie, alors que le temps écoulé depuis la fin des études diminue les chances d'endogamie. D'autre part, les chances d'hypergamie augmentent nettement avec l'âge (quoique moins fortement chez les femmes nées avant la Seconde Guerre mondiale), reflétant assez fidèlement l'évolution avec l'âge de la structure de la population des célibataires du sexe opposé. Au total, le rôle des fréquentations liées au cadre scolaire, et plus largement de la structure du marché conjugal dans lequel évolue l'individu, se trouve donc largement confirmé. Nous concluons par une mise en perspective de ces résultats avec ceux établis dans les chapitres précédents.

1 Contexte biographique de la mise en couple et choix du conjoint : théorie et littérature existante

Nous commençons par présenter les approches retenues par les travaux existants sur l'analyse biographique du choix du conjoint, tout en mobilisant des travaux portant sur la France pour affiner les hypothèses générales issues de cette littérature. Nous étudions successivement les deux dimensions qui nous semblent essentielles pour la définition du contexte de mise en couple : l'âge et la situation par rapport aux études et à l'emploi.

1.1 L'âge de mise en couple

1.1.1 Âge et contraintes structurelles sur le marché conjugal

L'âge d'un individu est à la fois un indicateur indirect de sa situation biographique – en particulier la situation scolaire et professionnelle, participant au passage à l'âge adulte –, et de sa position sur le marché conjugal. C'est ce second aspect qui nous intéresse ici, le premier étant étudié dans la section suivante. Il tient à un mécanisme simple : les individus choisissant généralement leur conjoint dans des tranches d'âge proches de la leur¹, l'avancée en âge réduit peu à peu le nombre de partenaires potentiels du fait des mises en couple qui ont déjà eu lieu au sein des cohortes de naissance proches. Plus un individu a attendu avant

1 Dans un quart à un tiers des couples, les conjoints sont nés deux années consécutives ; les cas dans lesquels l'écart d'âge dépasse 10 ans représentent moins de 10 % des couples, pourcentage en progression au fil des cohortes (Vanderschelden, 2006a).

de se mettre en couple, et plus son choix sera influencé par des contraintes structurelles liées à cette évolution (Oppenheimer, 1988 ; Smeenk, 1998, p. 40, 170). Selon certains travaux, les individus se mettant en couple de manière tardive seront poussés à réaliser un choix atypique sur le marché conjugal, et notamment à faire des choix hétérogames ou allant à l'encontre de la norme d'hypergamie féminine (Lichter, 1990).

Mais il nous semble essentiel de remarquer que les contraintes structurelles exercent leur influence dans un sens particulier : elles ne favorisent pas tous les choix « atypiques » de manière indifférenciée². En effet, au sein d'une cohorte de naissance donnée, la structure par diplôme de la population des célibataires évolue de manière assez marquée avec l'âge. De manière générale, ce sont les individus les plus diplômés qui se mettent en couple le plus tard : on pourrait donc penser que leur proportion parmi les célibataires au sein d'une cohorte va croissant avec l'âge. Néanmoins, un deuxième paramètre à prendre en compte³ est la proportion de célibataires définitifs dans chaque groupe de diplôme, qui joue principalement sur la composition de la population des célibataires à des âges élevés. Il reste donc à déterminer si l'interaction entre ces deux paramètres aboutit à une augmentation ou à une baisse du niveau de diplôme moyen des célibataires avec l'âge.

Concernant l'âge médian de première mise en couple, nous avons vu au chapitre précédent qu'il augmentait nettement avec le diplôme chez les femmes, mais moins nettement chez les hommes du fait de l'âge élevé de mise en couple des non diplômés – en particulier pour les anciennes cohortes. Concernant la proportion de célibataires définitifs, elle augmentait avec le diplôme chez les femmes des anciennes cohortes, mais aucune différence ne s'observe pour les cohortes récentes ; chez les hommes, cette proportion baisse au contraire avec le diplôme, et cette différence a persisté au fil des cohortes. Au total, on peut donc prédire avec une certitude assez forte que le niveau de diplôme des célibataires augmentera nettement avec l'âge chez les femmes, notamment au sein des anciennes cohortes : les deux paramètres agissent dans la même direction. Du côté des hommes,

2 En cela, l'évolution avec l'âge de la structure du marché conjugal se distingue de l'effet d'un *marriage squeeze* lié à la variation de la taille des cohortes de naissance (Muhsam, 1974), dont l'action principale n'est pas de modifier la composition de la population des célibataires, mais sa taille relative chez un sexe par rapport à l'autre.

3 À strictement parler, un troisième paramètre entre en jeu : la forme de la distribution des âges de mise en couple (troisième paramètre du modèle de Brüderl et Diekmann utilisé au chapitre précédent). L'interprétation fondée sur deux paramètres semble cependant suffisante pour résumer la plupart des tendances à l'œuvre.

l'évolution est difficile à prédire précisément du fait du lien ambigu entre âge de mise en couple et diplôme : on peut s'attendre à observer une certaine augmentation du niveau de diplôme moyen des célibataires aux âges faibles, mais qui sera progressivement compensée par l'augmentation de la part des célibataires définitifs, parmi lesquels sont sur-représentés les non diplômés. Ce renversement de tendance sera particulièrement fort pour les cohortes récentes, chez qui l'âge de mise en couple suit plus clairement le niveau de diplôme, et chez qui le taux de célibat définitif est particulièrement élevé chez les peu diplômés.

1.1.2 Des effets différents de l'âge selon le sexe ?

Ainsi, si la structure de la population des célibataires d'un âge proche a un effet sur les rencontres, comme nous en faisons l'hypothèse, les chances de rencontrer un conjoint très diplômé évolueront de manière différente avec l'âge chez les hommes et les femmes. Chez les premiers, celles-ci augmenteront nettement, reflétant l'évolution de la population des femmes célibataires. Chez les secondes, au contraire, et pour des raisons symétriques, l'augmentation sera moins nette. Si cette prédiction est vérifiée, un homme se mettant en couple relativement tard par rapport à son groupe de diplôme aura tendance, du seul fait des opportunités de rencontre auxquelles il est exposé, à choisir des conjointes plus diplômées que la moyenne de son groupe, et donc à effectuer un choix hypergame⁴. Une femme dans la même situation aura aussi des chances d'hypergamie plus élevées que la moyenne, mais moins nettement. Dans ce qui suit, nous mettrons conjointement à l'épreuve les hypothèses concernant l'évolution de la population des célibataires, et celles concernant les chances de rencontre proprement dites.

Ces facteurs structurels interagissent avec les normes sociales et les préférences individuelles (normes intériorisées), si bien que les chances de mise en couple hypergame peuvent éventuellement dévier de diverses manières du modèle proposé ci-dessus. En effet, les contraintes structurelles n'ont pas d'effet mécanique. Les individus se mettant tardivement en couple disposent de deux échappatoires (en plus du célibat) : choisir un conjoint nettement plus jeune⁵, ou un conjoint ayant déjà vécu en couple. Ces possibilités ne sont pas ouvertes de manière identique aux hommes et aux femmes, et les incitations à

4 La modélisation retenue plus bas, fondée sur le niveau absolu de diplôme du conjoint, ne distingue pas une augmentation de l'hypergamie d'une diminution de l'hypogamie. Nous parlerons par simplicité de l'hypergamie pour désigner en même temps, en négatif, l'hypogamie.

5 Choisir un conjoint au sein d'une cohorte plus âgée est bien entendu possible, mais la contrainte structurelle est dans ce cas encore plus forte que dans la cohorte de naissance de l'individu.

contourner les contraintes structurelles différent. La première des motivations qui pourraient aller à l'encontre d'une augmentation de l'hypergamie avec l'âge est la norme d'hypergamie féminine, qui pousserait les hommes à choisir des femmes plus jeunes pour éviter de se retrouver en situation d'infériorité par rapport à leur conjointe du point de vue du niveau d'éducation. Nous avons montré au chapitre précédent à quel point cette norme d'hypergamie féminine semblait fragile, puisqu'elle s'est immédiatement inversée lorsque les femmes sont devenues plus diplômées que les hommes. Il paraît donc douteux qu'elle suffise, même parmi les générations nées avant guerre, à contredire les contraintes structurelles du marché conjugal. Cette question constitue donc un test supplémentaire de l'existence d'une telle norme.

C'est particulièrement du côté des hommes que ce test retiendra notre attention, d'abord parce que nous nous attendons à une augmentation de l'hypergamie plus forte que chez les femmes, mais aussi du fait de la tendance à un écart d'âge en faveur de l'homme. Cet écart s'explique principalement par la préférence des jeunes femmes pour des partenaires plus âgés (Bozon, 1990a, 1990b ; Bergström, 2014, chap. 6). En effet, ces dernières valorisent avant tout chez les hommes la maturité sociale, qui s'acquiert progressivement avec l'obtention d'un diplôme, d'un emploi stable, puis l'avancement dans la carrière professionnelle (Bozon, 1990a ; Bihl et Pfefferkorn, 2002, chap. 4) : la mise en couple avec un conjoint « mûr » représente une voie privilégiée d'accession à l'indépendance pour les moins dotées scolairement (Bozon, 1990b). Ainsi, avec l'âge de mise en couple, l'écart d'âge en faveur de l'homme augmente, alors qu'il diminue chez les femmes : si les hommes, en vieillissant, prennent des conjointes de plus en plus jeunes qu'eux, les femmes ont tendance à choisir de plus en plus des conjoints d'un âge proche du leur (Vanderschelden, 2006a ; Mignot, 2010, p. 304), qui ont souvent déjà vécu en couple, ce qui est rendu possible par la plus forte tendance de ces derniers à se remettre en couple après une séparation (Cassan, Mazuy et Clanché, 2005). Du fait de ce phénomène, on peut considérer que les hommes se retrouvent en vieillissant en position de plus en plus favorable sur le marché conjugal (Bozon, 1990a, 1990b ; Ní Bhrolcháin et Sigle-Rushton, 2005) : leurs choix sont moins contraints que ceux des femmes puisqu'ils peuvent choisir des conjointes aussi bien du même âge que plus jeunes qu'eux⁶. Au total, si les hommes voient leurs chances de réaliser un choix hypergame

6 Cette tendance s'inverse cependant lentement après 30 ans, les femmes préférant au-delà de cet âge des conjoints d'un âge plus proche du leur (Bergström, 2014, p. 290)

augmenter avec l'âge, alors que leur position favorable leur permettrait de choisir au contraire des conjointes plus jeunes et moins diplômées, la réalité de la norme d'hypergamie féminine sera encore une fois battue en brèche.

L'évolution de l'hypergamie avec l'âge n'est pas moins intéressante du côté des femmes. Pour ces dernières, on l'a dit, la possibilité de choisir des conjoints nettement plus jeunes, n'est pas ouverte, ou de manière assez limitée : on pourrait donc s'attendre à ce que les contraintes structurelles jouent de manière importante, entraînant une nette hausse de l'hypergamie avec l'âge. Néanmoins, le modèle fortement genré suppose une difficulté de plus en plus grande à se mettre en couple pour les femmes avec l'avancée en âge : les hommes sont censés valoriser chez les femmes des caractéristiques relationnelles et esthétiques, ou encore un capital culturel, qui se développent peu avec l'âge. Ces difficultés, alliées au fait que les hommes ont la possibilité de choisir des conjointes plus jeunes qu'eux, ne seraient guère favorables à la rencontre d'un « bon parti », c'est-à-dire à la réalisation d'un choix hypergame, supposé être valorisé par les femmes. Prenons ici nos distances avec une lecture trop fréquemment caricaturale de ces différences genrées. Il ne paraît pas réaliste de considérer que les qualités esthétiques et la fécondité potentielle des femmes suffisent à définir leur valeur aux yeux des hommes, ni que ces critères subissent une érosion significative aux âges typiques de mise en couple⁷. Par conséquent, même si ce mécanisme recouvre probablement une certaine réalité, il nous paraît douteux d'affirmer que les femmes voient leur désirabilité véritablement diminuer avec l'âge sur le marché conjugal.

Notons par ailleurs que mesurer la désirabilité des femmes sur le marché conjugal à l'aune de leur capacité à se mettre en couple de manière hypergame révèle une difficulté conceptuelle concernant la mesure de la désirabilité des hommes. Se plaçant dans le cadre d'un modèle fortement genré, doit-on supposer, par symétrie avec les femmes, que les hommes les plus désirables sont ceux qui réalisent des choix hypogames, et prennent donc pour conjointes des femmes peu diplômées ? Il semble que ce serait pousser trop loin la

(Bergström, 2014, p. 290), et les mises en couple avec des femmes beaucoup plus jeunes restant peu fréquentes.

⁷ Soulignons par exemple que la période de fertilité des femmes suffit largement à donner naissance au nombre d'enfants typiquement observé dans les cohortes étudiées, même pour des premières mises en couple relativement tardives : moins de 20 % des individus ont eu plus de trois enfants dans toutes les générations considérées (Masson, 2013), et moins d'un quart des femmes se sont mises en couple après 26 ans parmi les cohortes nées avant les années 1960 (Prioux, 2003, p. 627). Voir aussi la critique de Melvyn G. Coles et Marco Francesconi (2011, p. 828), fondée sur des modèles de simulation de mise en couple.

lecture en termes de rôles de genre. Nous verrons plus bas que la similarité des mécanismes déterminant le choix du conjoint chez les hommes et les femmes incite plutôt à rejeter cette opposition radicale entre les deux sexes.

Au total, nous nous attendons à ce que l'effet des contraintes structurelles implique une augmentation avec l'âge des chances de rencontrer un conjoint très diplômé, et donc de réaliser un choix hypergame, une fois contrôlé notamment le diplôme finalement obtenu par l'individu. Ce mécanisme doit jouer surtout chez les hommes, et moins nettement chez les femmes. Chez les premiers, la norme d'hypergamie féminine, dont la réalité a déjà été mise en doute plus haut, ne nous semble pas en mesure d'aller véritablement à l'encontre des contraintes structurelles. Du côté des femmes non plus, même dans un modèle fortement genré, l'hypothèse d'une baisse avec l'âge du niveau moyen de diplôme des conjoints du fait d'une moindre désirabilité ne nous paraît pas réaliste. Nous montrerons en effet que nos deux hypothèses sont soutenues par les données empiriques.

Pour finir, notons que l'âge n'est pas la seule approche possible du passage du temps et de ses effets sur la position des individus sur le marché conjugal. Pris comme indice du contexte de la mise en couple, il présente l'inconvénient de mettre en équivalence des périodes de la vie dont le sens diffère profondément selon les milieux sociaux, particulièrement au sein des générations antérieures à la massification scolaire (Bourdieu, 1984 ; Galland, [1991] 2011). Le temps écoulé depuis la fin des études sera donc utilisé dans un second temps comme un indicateur alternatif du contexte de la mise en couple dans la trajectoire de l'individu. À diplôme contrôlé, nous pouvons nous attendre à observer les mêmes évolutions que suivant l'âge ; prendre comme référence la fin des études permet cependant de soumettre à une seconde épreuve l'interprétation proposée, et éventuellement de la préciser.

1.2 La scolarisation

1.2.1 L'influence des études sur les fréquentations

Second grand déterminant du contexte de mise en couple, la scolarisation peut agir de diverses manières sur le choix du conjoint. Comme nous le verrons plus bas, la massification scolaire et le développement des cohabitations hors mariage en ont fait un déterminant de plus en plus fort des modes d'entrée dans l'âge adulte et de la mise en couple. Pour analyser

l'effet de ce facteur, la démarche initiée par Robert Mare (1991) et développée par Hans-Peter Blossfeld, Andreas Timm et Faith Dasko (1997 ; Blossfeld et Timm, 2003b, p. 4), considérant le système scolaire comme un marché conjugal local, nous semble féconde – même si nous formulerons plus loin des réserves méthodologiques importantes concernant la manière de le mettre à l'épreuve des faits.

Cette approche part du constat (souvent désigné sous le nom d'« effet institution » des études), que le système scolaire regroupe des individus d'âge égal ou proche, et les fait progresser d'année en année vers un niveau supérieur, en sélectionnant à chaque étape une partie des élèves – plus ou moins importante selon les cohortes et les niveaux – qui est autorisée à poursuivre au niveau supérieur⁸. À mesure qu'un individu survit aux verdicts scolaires successifs et progresse vers un niveau supérieur, il fréquente des classes de plus en plus homogènes du point de vue du diplôme qui sera *in fine* obtenu par ses camarades. Illustrons cette évolution en comparant deux situations extrêmes : aux niveaux de scolarisation les plus bas, une classe offre, à la ségrégation résidentielle et scolaire près⁹, une approximation de la structure par diplôme finale de la cohorte considérée. Au contraire, au sommet de la hiérarchie scolaire, elle ne rassemble que deux groupes de diplôme, composés respectivement des élèves qui valideront et de ceux qui ne valideront pas leur année – une distinction qui sera elle-même souvent invisible dans des classifications relativement agrégées. Reprenant la tripartition des lieux de rencontre entre lieux ouverts, réservés et fermés établie par Michel Bozon et François Héran (1988), nous pouvons ainsi affirmer que le lieu d'études devient de plus en plus *réserve* à mesure que l'on s'élève dans le système scolaire.

Il suffit donc que les fréquentations scolaires déterminent au moins en partie les rencontres de conjoints potentiels pour que la scolarisation, et plus précisément la scolarisation à un niveau élevé, augmente les chances de former un couple endogame. Mais il semble nécessaire de compléter l'hypothèse retenue par la littérature : à strictement parler, ce sont les chances de réaliser un choix hypogame qui diminuent avec la progression dans le

8 Cf. Ugo Palheta (2011) ainsi que Mathieu Ichou et Louis-André Vallet (2013) pour des illustrations concrètes de ce processus dans le cas français. Notons qu'à une lecture purement verticale du système d'enseignement doit s'ajouter une analyse en termes de filières, lesquelles peuvent être rejointes à différents moments du cursus scolaire. Néanmoins, le résultat global reste le même, les classes devenant progressivement de plus en plus homogènes.

9 Cette ségrégation peut néanmoins être importante, même aux niveaux inférieurs, du fait en particulier des divisions précoces entre filières (voir la note précédente).

système scolaire, et celles de former un couple endogame *ou* hypergame qui augmentent. C'est que, paradoxalement, alors que l'élévation du niveau scolaire d'un individu rend de plus en plus improbable qu'il réalise un choix hypergame du fait de la diminution de la proportion de personnes plus diplômées que lui dans la population générale, sa progression dans le système scolaire le met en contact avec des individus dont le diplôme final sera de plus en plus fréquemment supérieur ou égal au sien. Ainsi, à *diplôme final contrôlé*, on doit s'attendre à ce que le temps passé en scolarisation avant la mise en couple augmente les chances d'hypergamie autant que celles d'endogamie.

1.2.2 Des fréquentations scolaires aux rencontres

Il semble inévitable que la scolarisation affecte fortement les opportunités de rencontres amoureuses et *in fine* de mise en couple. Le cadre scolaire fournit une partie importante des relations des jeunes scolarisés : les camarades d'études représentent plus de la moitié du réseau amical des 15-24 ans, et près des trois quarts pour ceux qui sont en cours d'études (Pan Ké Shon, 1998). Plus encore, les étudiants ont souvent tendance à fréquenter des personnes suivant le même cursus qu'eux (Erlich, 1998, chap. 8). Le style de vie lycéen ou étudiant est en effet propre à favoriser les rencontres avec des individus partageant le même statut, et à renforcer la coupure avec les individus de la même classe d'âge moins diplômés, déjà entrés sur le marché du travail, portés à se considérer comme plus « mûrs » (Bidart et Pellissier, 2002, p. 25)¹⁰. Et les relations formées dans une classe antérieure ne résistent souvent pas à l'orientation dans des filières différentes, pour des raisons à la fois d'éloignement physique et de centres d'intérêt – soit le sentiment de ne plus partager le même « univers » (Bidart et Pellissier, 2002, p. 40).

Or, si tous les cadres ne sont pas également favorables à des rencontres amoureuses, les fréquentations liées au cadre scolaire sont particulièrement adaptées à cette fonction. Du fait de la tendance des groupes de sociabilité étudiants à traverser les cercles, l'influence des fréquentations scolaires s'étend à l'ensemble des réseaux des individus, au-delà des contacts liés au cadre scolaire *stricto sensu*. En effet, la sociabilité lycéenne et étudiante se distingue par son organisation autour de « groupes » souvent constitués en partie de camarades d'une même classe (Bidart et Pellissier, 2002, p. 34), et qui donnent lieu à de nombreuses relations amoureuses (Bidart et Pellissier, 2002, p. 36, 41-43 ; Maillochon, 2003). Le rôle des soirées

¹⁰ Les termes repris entre guillemets sont ceux utilisés par les enquêtés des auteurs cités.

lycéennes et étudiantes ne paraît pas négligeable dans l'efficacité de ces réseaux à donner naissance à des couples (Juhem, 1995, p. 35 ; Erlich, 1998, p. 166) : M. Bozon et F. Héran ont souligné le rôle de la danse (et notamment celle *en couple*) comme rituel facilitateur de rencontres, à la fois dans le bal et dans les soirées entre amis, dont la part dans les lieux de rencontre progresse continûment depuis 1945 (Bozon et Héran, 1987, p. 951, 1988, p. 128).

De fait, les rencontres sur le lieu d'études représentent chez les deux sexes 25 % des premiers couples des diplômés du supérieur formés entre 1984 et 2006, contre moins de 10 % des rencontres des moins diplômés (Bozon et Rault, 2012, p. 468). Celles réalisées à l'occasion de fêtes entre amis, sans doute largement influencées par les réseaux amicaux liés aux études, représentent en outre entre 15 % et 20 % des premiers couples selon le niveau de diplôme des individus. Ces valeurs étaient semble-t-il plus faibles pour les générations antérieures : parmi les diplômés du supérieur, seuls 15 % des hommes et 13 % des femmes déclaraient avoir rencontré leur conjoint sur leur lieu d'études dans l'enquête Formation des couples de 1983-1984 (Bozon et Héran, 1988, p. 137) ; les fêtes entre amis gagnent nettement en importance avec le temps, passant de 2 % à 10 % des rencontres.

Enfin, outre que les fréquentations étudiantes constituent un vivier important où rencontrer son futur conjoint, les conseils des amis peuvent avoir une influence sur la stabilisation d'une relation ou son interruption (Bidart, 2008, p. 577). Et même l'opinion de simples connaissances peut avoir son importance, dans la mesure où la scolarisation entraîne une impossibilité d'échapper, le temps d'une année au moins, à l'« arène » de la classe, au sein de laquelle les jugements concernant l'assortiment des couples sont omniprésents (Juhem, 1995, p. 34). Tout semble donc concourir à ce que la scolarisation affecte le choix du conjoint¹¹.

Ces descriptions des réseaux amicaux des jeunes manquent cependant un point important pour les cohortes les plus anciennes : la mixité ne s'est généralisée dans l'enseignement secondaire français que dans les années 1960 et 1970¹². Les rencontres entre

11 Même si nous n'avons pas les moyens de creuser ici cette piste, il semble intéressant de souligner que l'organisation des classes préparatoires et des grandes écoles, plus proche que celle de l'université du modèle décrit ici, semble bien plus propice aux rencontres dans le cadre scolaire, ce qui est pleinement avec nos résultats présentés en première partie concernant la forte endogamie des diplômés des grandes écoles. Valérie Erlich note que la proportion des fréquentations au sein du même cursus est la plus élevée parmi les étudiants de classes préparatoires, de grandes écoles, d'IUT et de médecine (Erlich, 1998, p. 153).

12 Ce détail est curieusement ignoré aussi bien par H.-P. Blossfeld et A. Timm (2003b) que par Dominique Goux et Éric Maurin (2003) dans leur chapitre sur la France.

garçons et filles directement liées à la classe voire à l'établissement fréquentés étaient donc exclues au lycée pour les générations nées avant guerre. Cela n'empêche pas la scolarisation d'avoir un effet sur la mise en couple, étant donné, comme on l'a souligné, que les fréquentations lycéennes dépassent rapidement le cadre de la classe, et que la mixité existait à l'université. Ainsi, si l'hypothèse selon laquelle la scolarisation affecte le choix du conjoint paraît tout à fait crédible, il ne serait néanmoins pas étonnant de constater que cette influence était plus faible parmi les générations ayant précédé la généralisation de la mixité.

1.2.3 Les effets de la sortie du système scolaire

Finalement, si la scolarisation affecte les fréquentations des individus, et ce d'autant plus que leurs études ont duré longtemps, quel est l'effet de la sortie du système scolaire ? Avec l'entrée dans le monde du travail, les réseaux se recomposent rapidement (Bidart et Pellissier, 2002 ; Bidart et Lavenue, 2005, p. 365 ; et Bès, 2013 concernant spécifiquement les anciens d'une grande école d'ingénieurs) : suite à un éloignement physique et à un changement de styles de vie, les anciens camarades sont pour la plupart perdus de vue en quelques années, et seuls les liens les plus forts résistent. Les liens amicaux formés sur le lieu de travail prennent une part croissante avec l'âge, alors que ceux formés pendant les études deviennent plus rares, pour devenir moins nombreux que les premiers chez les trentenaires (Pan Ké Shon, 1998). Que l'insertion professionnelle soit rapide ou non, le nombre global de contacts diminue nettement – ce qui ne signifie pas que de nouveaux liens ne s'établissent pas, et aillent parfois au-delà des simples relations entre collègues, donnant accès à de nouveaux réseaux amicaux (Bidart et Pellissier, 2002, p. 30).

Or, on peut s'attendre à ce que les réseaux constitués dans ce cadre soient – dans une certaine mesure – moins homogènes du point de vue du diplôme que ceux liés aux études. Là où le système scolaire est organisé en classes d'élèves réunis précisément du fait de leur similarité pour suivre le même enseignement, l'organisation du travail entraîne plus de contacts entre des métiers différents. Mais surtout, du fait de la diversité des âges, qui contraste nettement avec le milieu scolaire, les collègues de travail représentent moins souvent des conjoints potentiels. Les célibataires actifs doivent donc recourir à d'autres lieux pour faire des rencontres – en particulier des lieux *ouverts* (Bozon et Héran, 1988) dont l'homogénéité du point de vue du diplôme ne peut être que plus faible que celle du lieu d'études. Ainsi, à mesure qu'augmente le temps depuis la fin des études, on peut donc

s'attendre à une plus grande hétérogénéité des rencontres du point de vue du diplôme, et donc à une moindre endogamie (Mare, 1991). En revanche, l'hypergamie, que la scolarisation est censée renforcer, doit aussi augmenter avec l'âge, selon les hypothèses avancées dans la section précédente : lequel des deux effets l'emportera reste une question ouverte.

Résumons les hypothèses que nous venons d'élaborer. Concernant tout d'abord l'effet de l'âge des individus, nous nous attendons à ce qu'il augmente, à diplôme final contrôlé, les chances de réaliser un choix hypergame. Plus précisément, l'existence d'une telle évolution chez les hommes serait une preuve supplémentaire de la fragilité de la norme d'hypergamie féminine du point de vue du diplôme. Chez les femmes, elle serait le signe que leur perte de désirabilité avec l'âge sur le marché conjugal est au mieux une exagération. Concernant ensuite l'effet de la scolarisation, nous avons suggéré qu'elle doit augmenter chez les deux sexes à la fois les chances d'endogamie et d'hypergamie, c'est-à-dire diminuer celles d'hypogamie. Le temps passé dans le système scolaire avant la mise en couple doit renforcer cet effet, alors que celui écoulé depuis la fin des études doit au contraire diminuer les chances d'endogamie – son effet sur l'hypergamie restant incertain.

1.3 Les travaux existants

La rareté des études biographiques du choix du conjoint ne nous permet pas d'avancer de nombreux éléments empiriques concernant la validité des hypothèses que nous venons d'avancer. Il paraît néanmoins utile de présenter rapidement les travaux existants ; nous reviendrons plus en détail sur leurs résultats au moment de discuter les nôtres.

Le premier, Daniel T. Lichter (1990) a tenté de mesurer l'effet d'un mariage tardif sur le choix du conjoint des femmes en termes d'âge, d'appartenance ethnique, d'histoire matrimoniale et de diplôme. Mais le modèle retenu ne contrôlait pas le fait que les plus diplômées se mettent en couple le plus tardivement, ce qui introduit un effet de sélection majeur. Portant l'attention sur l'effet du système scolaire comme marché conjugal, R. Mare (1991) a montré, à l'aide de modèles log-linéaires portant sur des cohortes de mariages aux États-Unis, que le temps écoulé entre la fin des études et la mise en couple diminuait l'endogamie d'éducation. Cette approche a été reprise dans un ouvrage collectif par H.-P.

Blossfeld et A. Timm (2003a) sur le modèle d'une étude antérieure (Blossfeld, Timm et Dasko, 1997), mais suivant une méthodologie qui nous paraît incorrecte, et sur laquelle nous reviendrons plus bas. Wilma Smeenk (1998) a poursuivi cette piste, mais aussi celle des effets de l'âge, et a notamment observé que la scolarisation et l'avancée en âge diminuaient les chances d'hypogamie, c'est-à-dire augmentaient celles d'hypergamie et/ou d'endogamie (quoique moins clairement chez les femmes que chez les hommes). Malheureusement, la faible taille des échantillons utilisés a probablement empêché la mesure d'autres effets plus fins. Finalement, en France, Mélanie Vanderschelden (2006b) a confirmé, à l'aide d'une régression logistique à partir de l'enquête Étude de l'histoire familiale 1999, que l'endogamie de diplôme était plus forte parmi les couples qui s'étaient formés pendant ou peu après les études.

L'apport de la présente étude est de combiner l'analyse des effets de l'âge et de la scolarisation, tout en modélisant les chances de rencontrer un conjoint de chaque niveau de diplôme, et pas seulement les chances d'endogamie. Mais aussi de mettre en évidence des évolutions temporelles, ce qui n'avait à notre connaissance jamais été fait. Ces deux avancées sont rendues possibles par la taille exceptionnelle de l'échantillon de l'enquête Étude de l'histoire familiale.

2 Mise en couple et scolarisation dans les enquêtes Étude de l'histoire familiale et Biographies et entourage

Nous recourons ici encore aux enquêtes Étude de l'histoire familiale (Insee-Ined, 1999) et Biographies et entourage (Ined, 2001) : nous renvoyons donc le lecteur au chapitre précédent pour des informations générales concernant ces enquêtes. Nous abordons dans cette section une question centrale pour la présente analyse et qui n'a pas été traitée précédemment : la datation conjointe de la mise en couple et de la fin des études d'Ego.

2.1 Étude de l'histoire familiale

L'enquête Étude de l'histoire familiale renseigne l'année et le mois de naissance et de mise en couple¹³ (cf. chapitre précédent), mais l'âge de fin d'études est déclaré en années : il

13 Rappelons ici que la vie en couple est définie comme une cohabitation ayant duré plus de 6 mois. Nous ne pouvons cependant exclure que certains enquêtés aient trouvé plus naturel de renseigner la date de début

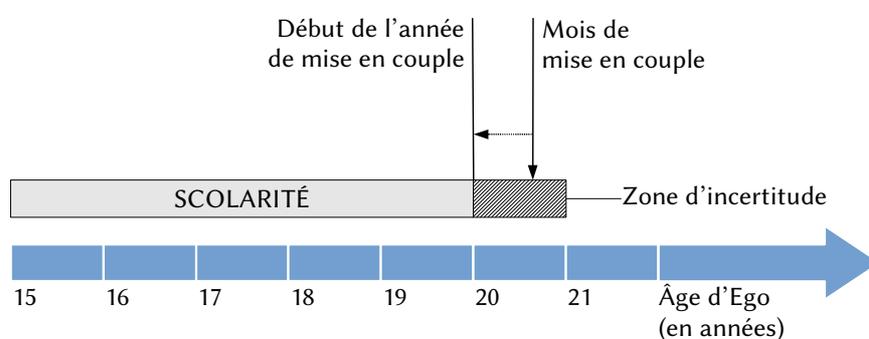
ne nous est donc pas possible d'identifier précisément le mois d'interruption des études. Un tel niveau de précision ne paraît de toute façon pas essentiel, étant donné que les processus considérés – fréquentations dans le cadre scolaire, début de relation, puis décision de cohabiter – se déroulent plus à l'échelle d'une année que d'un mois : la notion d'un « temps flou » (Courgeau et Lelièvre, 1988, 1989, p. 82), nécessaire pour passer des décisions et des circonstances qui les entourent à la réalisation d'un événement, s'applique ici pleinement. Nous travaillons donc exclusivement en années, mesurées par l'âge de l'interrogé (que nous désignerons par la suite sous le nom d'Ego).

Puisque la déclaration de l'âge de fin d'études en années revient à réaliser une troncature à l'année inférieure, nous effectuons la même opération concernant l'âge de première mise en couple, de manière à rendre ces deux informations comparables. Ainsi, formellement, ces variables nous indiquent, pour l'une, qu'Ego était scolarisé à l'âge donné, mais qu'il ne l'était plus un an après ; et, pour l'autre, qu'il s'est mis en couple pour la première fois entre ces deux âges. Lorsque les deux événements ont lieu la même année, nous ignorons lequel des deux est arrivé en premier, et donc si la mise en couple a réellement eu lieu pendant la scolarité. Nous choisissons de trancher ce dilemme en définissant la période de scolarisation dans un sens extensif : nous considérons qu'Ego est scolarisé jusqu'à l'anniversaire suivant sa fin d'études (Figure 6.1). Ce choix se justifie par le fait que la date de mise en couple dont nous disposons correspond à l'entrée en cohabitation, et non à celle de la rencontre ou du début de la relation. Étant donné que nous nous intéressons aux déterminants du choix du conjoint, il est préférable de mesurer les facteurs légèrement en amont de la date d'entrée en cohabitation, plutôt qu'en aval de celle-ci, afin de tenir compte de la période de maturation de l'événement. Plus généralement, la logique impose que les facteurs explicatifs (ici, la scolarisation) prennent place avant les phénomènes qu'ils sont supposés influencer (le choix du conjoint et la mise en couple)¹⁴.

de relation, devenue cohabitante seulement plus tard.

14 Encore que cette évidence apparente soit plus complexe à manier qu'il n'y paraît dans le contexte d'une analyse statistique – point sur lequel nous reviendrons plus précisément lors de la discussion des modèles.

Figure 6.1 : Méthode de comparaison de l'âge de fin de scolarité et de l'âge de mise en couple lorsque les deux événements ont lieu la même année



La manière de traiter les cas dans lesquels scolarisation et mise en couple ont lieu la même année, sans que l'on sache lequel des deux événements a eu lieu le premier, n'a pas qu'une influence anecdotique sur les résultats. En effet, si la proportion des mises en couple qui ont eu lieu la même année que la fin des études est de seulement 2 % pour les cohortes nées dans les années 1920, elle passe à 15 % pour celles nées dans les années 1960 – du fait de la diminution du temps moyen écoulé entre ces deux événements, que nous présenterons plus bas. Ainsi, selon la définition – restrictive ou extensive – de la scolarité adoptée, la proportion de personnes s'étant mises en couple en cours de scolarité varie, toutes cohortes confondues, entre 7 et 17 %. Les proportions présentées dans la section suivante, fondées sur la définition extensive développée au précédent paragraphe, représentent donc une estimation haute, et doivent être interprétées comme une évaluation de l'importance des rencontres en cours d'études plus que de celles des entrées en cohabitation *stricto sensu*.

2.2 Biographies et entourage

Contrairement au chapitre précédent, nous ne serons pas en mesure de confirmer les analyses avec Biographies et entourage, du fait des trop faibles effectifs de cette enquête pour l'estimation de modèles complexes¹⁵. En revanche, elle nous sera utile pour mieux appréhender, à titre de première approche, l'effet de la scolarisation sur le choix du conjoint. En effet, contrairement à Étude de l'histoire familiale, Biographies et entourage nous

15 Les données de Biographies et entourage ne permettent presque jamais d'obtenir dans les analyses qui suivent des résultats significatifs, même avec des modèles simplifiés.

renseigne sur la date de première activité d'Ego, mais aussi sur celle de son premier conjoint.

Nous ne disposons ici que de l'âge en années, que ce soit concernant la mise en couple¹⁶ ou l'entrée en activité. La question de ramener le mois de mise en couple à l'anniversaire précédent ou suivant ne se pose donc pas : la troncature est réalisée par l'enquête. Notons que cette information est parfaitement cohérente avec les choix décrits dans la section précédente concernant Étude de l'histoire familiale.

Le calcul de l'âge de fin d'études est plus délicat. Contrairement à Étude de l'histoire familiale, Biographies et entourage ne renseigne pas séparément la date de fin d'études. Nous n'avons connaissance que de la première activité d'Ego et de celle de son conjoint après la fin de ses études, mais selon une définition plus large que le premier emploi : le service militaire, le chômage, l'invalidité ou encore le statut de personne au foyer sont inclus dans le champ. Concernant Ego, la première activité est simplement la première ligne renseignée dans la colonne activités de la fiche âge-événement (AGEVEN) utilisée pour décrire l'ensemble de la biographie d'Ego. Concernant son premier conjoint, dont la biographie est nécessairement moins détaillée, le questionnaire commence par interroger Ego sur l'activité de ce dernier lors de la mise en couple, en précisant explicitement la possibilité d'être étudiant. Est ensuite abordée, entre autres, la première activité du conjoint. Ainsi, si la date exacte de fin d'études n'est pas précisément connue, le statut lors de la mise en couple fait l'objet d'une question explicite. C'est justement cette information que nous utiliserons dans la section suivante.

En dépit de la définition large de la première activité retenue dans le questionnaire, il apparaît que certains individus n'ont pas déclaré d'activité jusqu'à un âge assez avancé. Cette absence d'information semble devoir s'interpréter comme le signe que les enquêtés (ou les enquêteurs) n'ont pas considéré que des périodes d'inactivité avant la mise en couple renaient dans la définition attendue. Nous avons donc imputé, pour les individus dont l'âge de première activité dépassait de plus de trois ans le quartile supérieur de leur groupe

16 Rappelons que la définition du couple retenue par cette enquête exige un an de vie commune, la date renseignée étant celle de la mise en ménage. L'enquête renseigne aussi, pour les couples qui n'ont jamais cohabité, la date de début de relation, qui serait d'un grand intérêt pour notre étude ; malheureusement, cette information n'a pas été demandée aux couples ayant finalement cohabité, soit la grande majorité d'entre eux (Lelièvre *et al.*, 2006, p. 40).

de sexe et de diplôme (soit 4 % des conjoints et 2 % des interrogés), l'âge médian observé au sein de ce groupe.

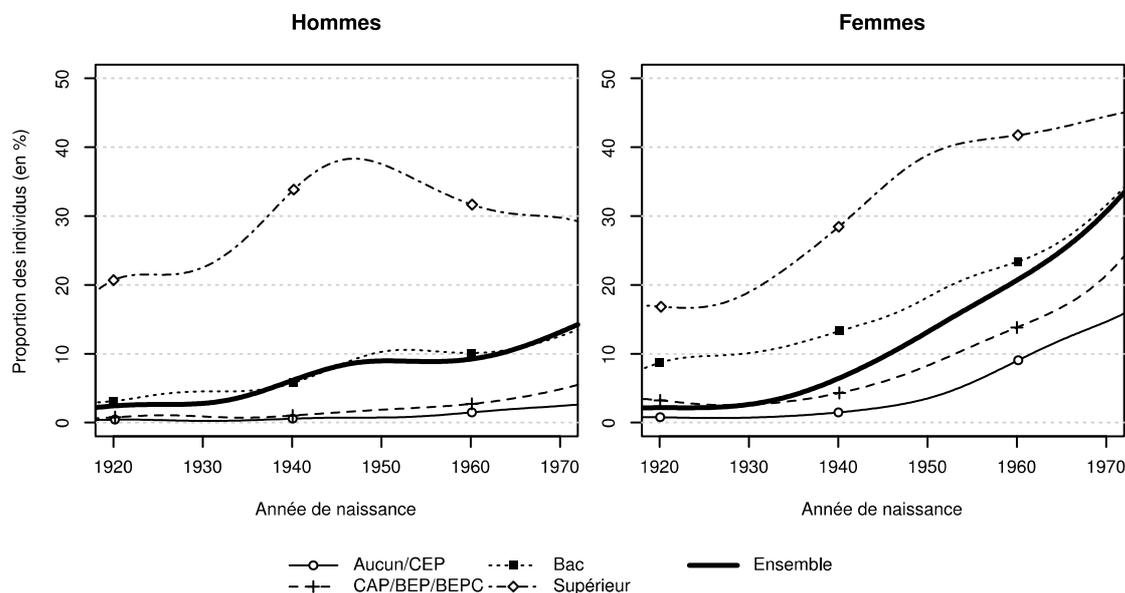
3 Mise en couple et scolarisation : un état des lieux

Avant d'étudier un potentiel effet de la scolarisation et du temps écoulé depuis celle-ci sur les chances de mise en couple avec un conjoint d'un niveau de diplôme donné, il semble essentiel de mesurer la fréquence de ces situations, et de décrire les couples formés dans ce contexte du point de vue des diplômes des deux conjoints.

3.1 Le développement des mises en couple pendant et peu après les études

Les mises en couple en cours d'études se sont nettement développées au fil des cohortes (Robert-Bobée et Mazuy, 2005, p. 181) : elles concernaient moins de 5 % des individus des deux sexes parmi les cohortes nées avant la Seconde Guerre mondiale, mais 15 % des hommes et 30 % des femmes parmi celles nées à la fin des années 1960 (Figure 6.2). Cette augmentation s'explique bien sûr en partie par la massification scolaire, mais elle s'observe aussi au sein de chaque groupe de diplôme. Il semble ainsi qu'elle doive beaucoup aux transformations du couple qui ont réduit l'incompatibilité entre statut d'étudiant et mise en ménage : cohabitation avant et même en dehors de tout projet de mariage, contraception. Rappelons néanmoins que notre définition est ici extensive, incluant des mises en couple qui ont pu avoir lieu dans l'année suivant la fin des études.

Figure 6.2 : Proportion des individus s'étant mis en couple pendant leurs études selon le diplôme et l'année de naissance



Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : individus nés en France

Une régression locale de degré 1 (lissage non paramétrique) est utilisée (cf. la légende de la figure 3.3, p. 202).

Les mises en couple en cours d'études sont ici rapportées à l'ensemble des individus, y compris ceux n'ayant jamais vécu en couple. Cette mesure permet d'éviter le biais lié au fait qu'une partie des individus d'une cohorte n'ont jamais vécu en couple à la date de l'enquête (censure à droite) ; en revanche, on peut raisonnablement faire l'hypothèse que l'ensemble des interrogés ont terminé leurs études en 1999 (soit à l'âge de 29 ans pour les individus nés en 1970).

Ce mouvement va de pair avec une augmentation de la proportion des rencontres du premier conjoint qui se font dans le cadre des études. Considérant une cohorte de premiers couples dans son ensemble, ce sont moins de 5 % des couples formés au début du XX^e siècle qui s'étaient rencontrés sur leur lieu d'études, et 2 % au cours d'une fête entre amis (Bozon et Héran, 1987, p. 950). Ces chiffres augmentent nettement au fil des cohortes, pour atteindre un peu moins de 10 % pour chacun de ces deux lieux au sein des couples formés dans les années 1980. Quelques années plus tard, ce sont 15 % des premiers couples formés entre 1984 et 2006 qui se sont rencontrés dans un cadre scolaire, et 18 % dans une soirée entre amis (Bozon et Rault, 2012, p. 459-460). Sans que l'on puisse inférer une identité entre ces rencontres et les couples formés en cours d'études que nous observons ici, nous pouvons

remarquer que l'ordre de grandeur des deux phénomènes est très similaire pour toutes les cohortes chez les hommes. Chez les femmes, la proportion de mises en couple en cours d'études augmente bien plus fortement que celle des rencontres dans ce cadre, indiquant un décalage entre ces deux phénomènes. On peut sans doute l'expliquer par la persistance d'un écart d'âge moyen en faveur de l'homme, qui implique que les femmes se mettent en couple en cours d'études avec des hommes qui les ont déjà terminées (Bozon, 1990b). Cette explication semble trouver une certaine confirmation (*cf. infra*) dans le fait que l'effet de la scolarisation sur l'endogamie paraît moins fort chez les femmes que chez les hommes.

De manière attendue, la mise en couple en cours d'études devient plus fréquente à mesure que l'on s'élève dans la hiérarchie des diplômes, ce qui correspond ici encore à la variation de la fréquence des rencontres dans le cadre scolaire déjà présentée plus haut. Chez les hommes, ce n'est que chez les diplômés du supérieur que ce phénomène prend une véritable importance. Au contraire, la progression est régulière chez les femmes. Les proportions assez élevées observées même chez les non diplômées incitent à penser que la définition extensive que nous avons retenue, ainsi sans doute qu'une interprétation elle aussi large du fait de vivre en couple, convergent pour fournir une estimation haute des mises en couple en cours de scolarité.

On peut noter une divergence importante entre hommes et femmes diplômés du supérieur parmi les cohortes du baby-boom : alors que chez ces dernières les mises en couple en cours de scolarité se sont développées de manière continue, on observe chez les premiers un pic suivi d'une baisse de 10 points de pourcentage. Cette inflexion s'explique sans doute par le retournement de la conjoncture économique et la hausse du chômage, qui rendent plus incertaines les perspectives d'emploi des étudiants¹⁷ : on retrouve ici un mécanisme déjà observé au chapitre précédent à propos de l'âge de mise en couple et du célibat des non diplômés. Ces nouveaux résultats apportent ainsi une confirmation de l'importance persistante de la stabilité professionnelle sur la mise en couple des hommes (Solaz, 2000), et peuvent expliquer que les mises en couple en cours d'études demeurent moins fréquentes chez les hommes.

17 *A contrario*, la prospérité économique est considérée par Oppenheimer (1988, p. 579-580) comme l'explication de la baisse de l'âge au mariage des cohortes de l'immédiat après-guerre.

Au total, il apparaît que la mise en couple avant la fin des études est un phénomène assez fréquent chez les femmes, ainsi que chez les hommes les plus diplômés, et surtout parmi les cohortes récentes. Mais au-delà, la tendance à se mettre en couple en cours d'études peut être considérée comme un cas limite relevant du phénomène plus général de la mise en couple peu de temps après la fin des études. L'interaction entre les âges de fin d'études et de première mise en couple des individus peut jouer sur le choix du premier conjoint, en particulier par le biais des rencontres liées au cadre scolaire.

3.2 Le rapprochement entre fin d'études et mise en couple

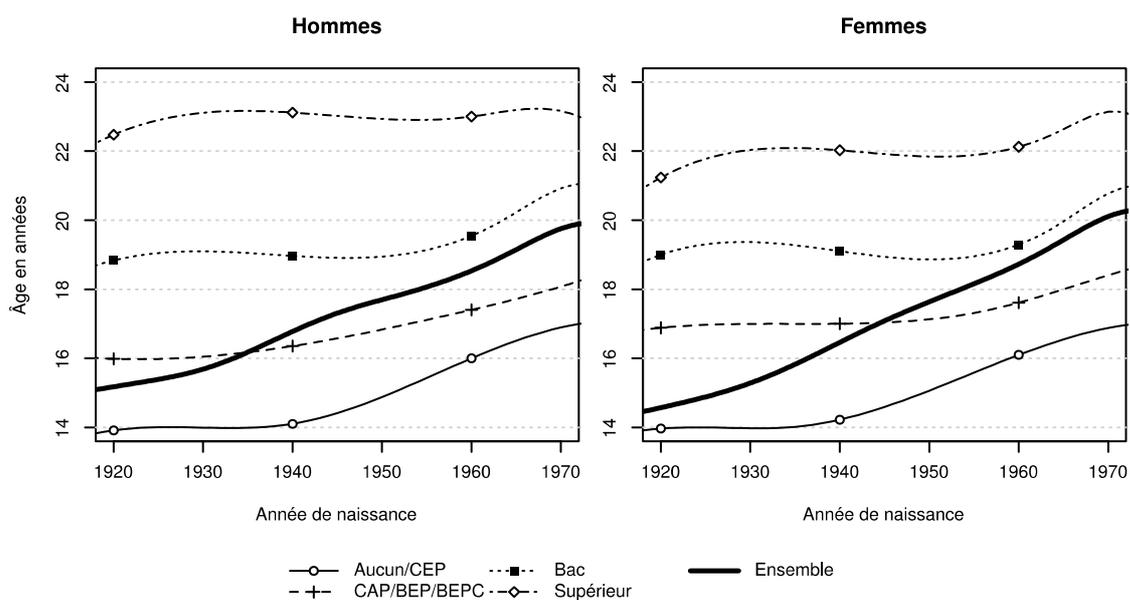
L'augmentation de l'âge médian de fin d'études lié à la massification scolaire constitue la principale raison de l'augmentation de la fréquence des mises en couple en cours d'études. Cette évolution correspond à la première explosion scolaire commencée avec les générations nées au milieu des années 1930 ; en revanche, nous n'observons pas la seconde explosion scolaire, qui ne démarre qu'avec les cohortes de la fin des années 1960 (Chauvel, 1998, p. 8). L'augmentation est visible aussi bien au niveau de l'ensemble d'une cohorte : l'âge médian de fin d'études passe de 14-15 ans à 20 ans entre la cohorte 1920 et la cohorte 1970 (Figure 6.3). Mais aussi au sein de chaque groupe de diplôme (à l'exception des diplômés du supérieur), du fait d'une part de l'augmentation de l'âge de fin de scolarité obligatoire¹⁸ (pour les non diplômés), et d'autre part du développement des abandons après une tentative de prolongation d'études à l'université (Beaud, 2002).

En comparaison de ce mouvement d'ampleur, les variations de l'âge médian de mise en couple, de l'ordre d'une à deux années, sont relativement faibles. Ses évolutions selon le diplôme ont déjà été commentées au chapitre précédent : rappelons simplement ici que l'on observe chez les deux sexes une évolution en U, la première mise en couple des baby-boomers ayant été la plus précoce du siècle. De ce fait, conformément à ce qu'avaient déjà noté France Prioux (2003, p. 633) ainsi qu'Isabelle Robert-Bobée et Magali Mazuy (2005, p. 181) pour la France, et R. Mare (1991) pour les États-Unis, l'écart entre la fin des études et la première mise en couple n'a cessé de se réduire au fil des cohortes (Figure 6.4), diminuant de 5 ans chez les deux sexes. Tous les niveaux de diplôme ont été affectés de manière similaire. La remontée de l'âge de première mise en couple depuis la cohorte 1955 ne semble

18 De 14 ans depuis la loi Zay de 1936 (s'appliquant aux enfants nés après 1923), celui-ci est porté à 16 ans par la loi Berthoin en 1959 (enfants nés après 1953).

avoir suffi à inverser le mouvement que chez les hommes diplômés du supérieur. Ainsi, si le contexte scolaire a des effets différents sur le choix du conjoint selon la cohorte de naissance, ce mécanisme doit agir autant par la proximité temporelle entre études et mise en couple, que par la mise en couple pendant les études *stricto sensu*.

Figure 6.3 : Âge médian de fin d'études selon le diplôme et l'année de naissance

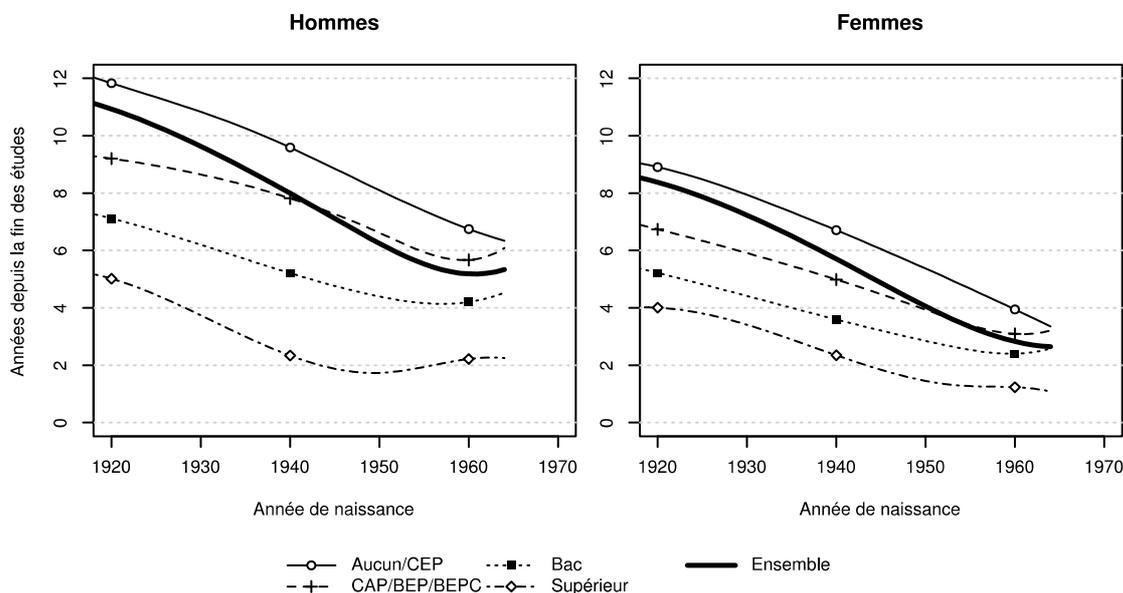


Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : individus nés en France

Une régression quantile comprenant des splines cubiques restreintes à 4 degrés de liberté est utilisée, ce qui revient à lisser la courbe de manière non paramétrique (*cf.* Encadré 5.2, p. 308).

Figure 6.4 : Temps médian écoulé entre la fin des études et la mise en couple par niveau de diplôme et année de naissance



Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : individus nés en France et ayant vécu en couple cohabitant

Note : sont inclus les individus ayant cohabité pendant leurs études (temps écoulé nul)

Une régression quantile comprenant des splines cubiques restreintes à 4 degrés de liberté est utilisée, ce qui revient à lisser la courbe de manière non paramétrique (*cf.* Encadré 5.2, p. 308). L'âge médian de mise en couple par diplôme n'est pas observable au-delà de la cohorte 1965 pour les groupes les plus diplômés.

3.3 Qui sont les couples formés en cours d'études ?

Nous venons de voir que la fréquence des mises en couple en cours d'études a nettement augmenté au fil des cohortes. Pour mieux cerner l'effet de cette évolution sur le choix du conjoint, et avant de passer à des analyses plus sophistiquées, il paraît essentiel de tenter de mieux cerner le profil de ces couples. Les données de Biographies et entourage sont ici précieuses en ce qu'elles nous permettent d'appréhender la situation des deux conjoints, et pas seulement celle des interrogés.

Dans la démarche retenue ici, nous nous intéressons en particulier aux diplômes finalement obtenus par les deux conjoints, à leur statut au moment de la mise en couple, ainsi qu'à leur âge. Ces trois dimensions sont intimement liées : par exemple, un individu ayant déjà terminé ses études pourra difficilement trouver un conjoint moins diplômé que lui qui soit encore scolarisé au sein de sa classe d'âge.

Tableau 6.1 : Statut d'activité des conjoints lors de la première mise en couple d'Ego

	Aucun scolarisé	Ego scolarisé	Conjoint scolarisé	Tous les deux scolarisés
Hommes	81	4	10	4
Femmes	80	11	5	3

Lecture : dans 4 % des premières mises en couple des hommes, ceux-ci étaient scolarisés alors que leur conjointe ne l'était plus.

Source : enquête Biographies et entourage (Ined)

Champ : premiers couples des individus nés en France entre 1930 et 1950 et résidant en Île-de-France en 2001

8 % des hommes et 14 % des femmes de Biographies et entourage se sont mis en couple pendant leurs études (Tableau 6.1), soit des proportions légèrement plus élevées que celles présentées plus haut pour les mêmes cohortes concernant la France entière à partir d'Étude de l'histoire familiale. On note une relative symétrie des premiers couples des hommes et des femmes¹⁹ : chez les deux sexes, environ 18 % des premiers couples se sont formés pendant la scolarisation d'au moins un des deux conjoints. Dans le cas, le plus fréquent, où un seul des deux conjoints est scolarisé lors de la mise en couple, c'est le plus souvent la femme qui est dans cette situation (environ 10 % des cas, contre environ 5 % pour les hommes).

Les cas dans lesquels les deux conjoints sont scolarisés sont pour leur part très rares (entre 3 et 4 % des cas). Néanmoins, on observe une association positive entre la scolarisation de l'un et de l'autre conjoint à la mise en couple : ainsi, 48 % des hommes et 22 % des femmes qui se sont mis en couple avant la fin de leurs études l'ont fait avec un conjoint lui-même encore scolarisé, contre respectivement 11 % et 6 % pour les hommes et les femmes qui se sont mis en couple après la fin de leurs études²⁰.

19 Symétrie due au fait que seuls 10 % des individus se mettent en couple avec une personne ayant déjà vécu une expérience conjugale.

20 Chez les deux sexes, cette association est hautement significative selon un test du Khi2. Par ailleurs, l'association positive entre scolarisation d'Ego et scolarisation de son premier conjoint lors de la mise en couple persiste lorsque l'on contrôle le diplôme des deux conjoints, voire leur écart d'âge, dans un modèle de régression logistique.

Même si la mise en couple en cours de scolarisation s'est fortement développée, le caractère fortement minoritaire de ces situations indique que le lien entre mise en couple en cours de scolarité et endogamie est loin d'être aussi évident que les éléments présentés plus haut porteraient à le croire : dans la plupart des cas où Ego est scolarisé à la mise en couple, le conjoint a déjà terminé ses études, ce qui indique qu'il ne suit pas (ou plus) le même parcours scolaire qu'Ego. Ce résultat n'exclut pas à lui seul que la rencontre ait été influencée par le cadre scolaire : nous observons ici la première cohabitation, nécessairement postérieure à la rencontre. Mais il soulève un point important : une rencontre sur le lieu d'études n'équivaut pas nécessairement à la formation d'un couple endogame ; il suffit que l'un des conjoints interrompe ses études avant l'autre pour qu'ils obtiennent *in fine* des diplômes différents.

Or, c'est bien ce qui semble avoir lieu dans les cas de scolarisation d'un seul des deux conjoints. En effet, deux scénarios sont alors envisageables : soit le conjoint est plus jeune ou du même âge qu'Ego, mais moins diplômé, ce qui explique qu'il ait terminé ses études avant Ego ; soit il est plus âgé, auquel cas il peut être indifféremment plus, moins ou également diplômé qu'Ego. Le premier cas exclut clairement l'endogamie ; le second ne la rend *a priori* ni plus, ni moins probable que si aucun des conjoints n'était scolarisé. L'étude de l'écart d'âge entre conjoints va nous permettre de savoir lequel de ces deux scénarios est le plus fréquent.

Tableau 6.2 : Écart d'âge en faveur de l'homme selon le statut d'activité des deux conjoints

	Aucun scolarisé	Ego scolarisé	Conjoint scolarisé	Tous les deux scolarisés
Hommes	1,9	-0,2	4,4	0,8
Femmes	2,8	4,6	-0,4	1,4

Lecture : les hommes qui étaient encore scolarisés lors de leur première mise en couple et dont la conjointe ne l'était plus sont en moyenne plus jeunes de 2 mois (0,2 années) que cette dernière.

Source : enquête Biographies et entourage (Ined)

Champ : premiers couples des individus nés en France entre 1930 et 1950 et résidant en Île-de-France en 2001

Le croisement de l'écart d'âge entre conjoints avec leur statut d'activité à la mise en couple répond clairement à notre interrogation (Tableau 6.2). Commençons d'abord par confirmer le caractère plausible du lien entre scolarisation des deux conjoints à la mise en couple et rencontre liée aux études. En effet, dans cette configuration, l'écart d'âge en faveur de l'homme est fortement réduit par rapport aux autres cas – même s'il ne disparaît pas complètement. Ceci indique que le choix du conjoint se porte alors sur des individus de la même classe d'âge, soit des fréquentations typiques de l'organisation du système scolaire. De fait, aussi bien M. Bozon et F. Héran (1987, p. 958) concernant la France que W. Smeenk (1998, chap. 4) à propos des Pays-Bas ont montré que les couples qui se sont rencontrés dans le cadre scolaire – sans nécessairement avoir cohabité à cette période – présentent un écart d'âge plus faible que la moyenne des couples²¹. Nous pouvons donc considérer un faible écart d'âge comme un indice d'une rencontre liée aux fréquentations scolaires, ce qui va tout à fait dans le sens du cadre théorique développé plus haut.

En revanche, lorsqu'un seul des deux conjoints est scolarisé à la mise en couple, l'écart d'âge moyen se déplace, par rapport à l'ensemble des premiers couples, en faveur du conjoint qui est déjà en activité²². Mais ce déplacement présente des effets opposés selon le sexe : il revient à réduire l'écart chez les hommes, et à l'augmenter chez les femmes. Les deux scénarios présentés ci-dessus sont donc valables, mais chacun pour un sexe. Les hommes scolarisés à la mise en couple et dont la conjointe ne l'est plus ont choisi des femmes en moyenne plus âgées qu'eux de quelques mois seulement, et donc nécessairement moins diplômées, étant donné qu'elles ont terminé leurs études avant eux (premier scénario). Les femmes dans la même situation ont au contraire choisi des hommes plus âgés qu'elles de cinq ans en moyenne (second scénario) : avec un tel écart d'âge, le fait qu'elles soient encore scolarisées, contrairement à leur conjoint, n'a pas d'implications claires en termes d'écart de diplôme entre les conjoints.

La scolarisation de l'homme, de la femme, ou des deux conjoints correspondent donc à trois configurations très différentes, qui ont en conséquence des effets assez contrastés sur l'écart entre diplômes des conjoints (Figure 6.5). On distingue tout d'abord, comme attendu,

21 De même, cette dernière auteure a observé (1998, p. 129-130) que les conjoints qui se sont rencontrés dans un cadre scolaire étaient moins souvent actifs que les autres lors de leur mise en couple, ce qui semble indiquer une corrélation entre rencontre en cours d'études et mise en couple dans ce même contexte.

22 Cet effet avait déjà été observé par Mélanie Vanderschelden (2006a, p. 3) et Jean-François Mignot (2010, p. 311), mais sans informations concernant la scolarisation du conjoint de l'interrogé.

une propension à l'endogamie nettement plus forte au sein des couples formés pendant la scolarisation des deux conjoints (quadrant supérieur gauche), qui atteint même 80 % parmi les premiers couples des femmes – soit 23 points de pourcentage de plus que lorsque aucun conjoint n'est scolarisé²³. Mais ces configurations représentent, comme on l'a vu, une proportion très faible des premiers couples, ce qui limite l'effet global de la scolarisation sur l'endogamie.

En revanche, la sur-représentation des couples endogames parmi les couples formés pendant la scolarisation d'un seul des deux conjoints est plus modeste : l'écart maximal par rapport à la situation dans laquelle aucun conjoint n'est scolarisé est de 12 points de pourcentage chez les femmes, et de 6 points seulement chez les hommes. C'est l'équilibre entre hypergamie et hypogamie qui se modifie nettement : le conjoint encore scolarisé lors de la mise en couple obtiendra bien plus souvent un diplôme supérieur à celui de son conjoint à l'issue de ses études que dans les autres configurations. Ce mouvement, combiné avec une tendance générale à l'hypergamie féminine, aboutit à une quasi-disparition des couples à hypergamie masculine lorsque l'homme est scolarisé et sa conjointe non ; et au contraire à un équilibre entre hypogamie et hypergamie lorsque l'inverse est vrai. La symétrie des premiers couples des hommes et des femmes est de ce point de vue frappante.

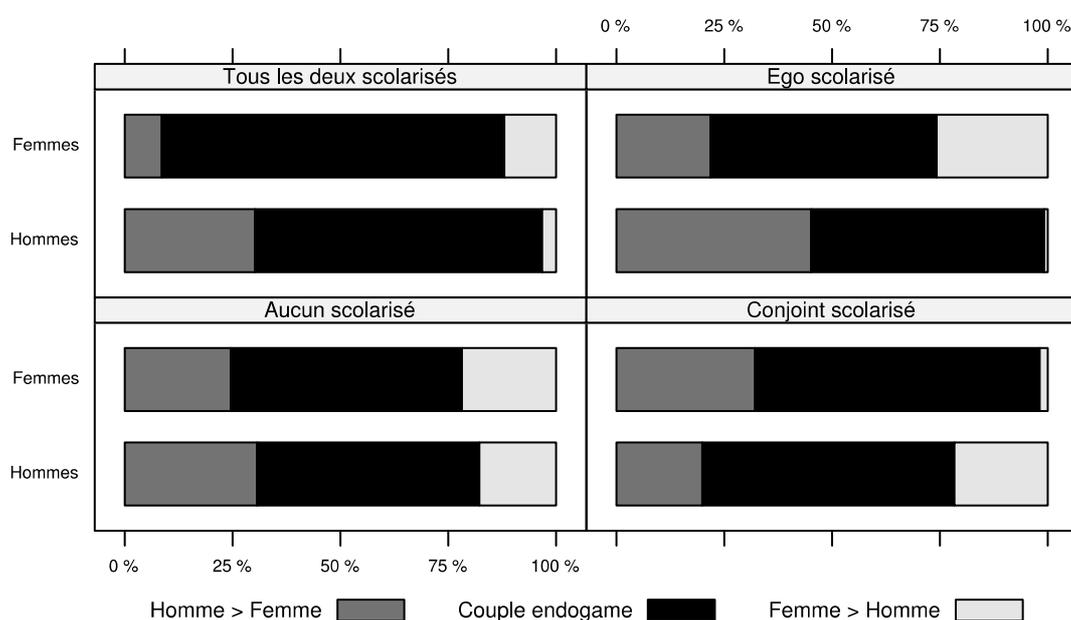
Prenons garde à ce stade à ne pas interpréter ce résultat comme l'indice du fait que la scolarisation renforcerait l'hypogamie, ce qui irait à l'encontre de l'hypothèse formulée plus haut. En effet, comme nous l'avons déjà souligné, les individus qui ont les scolarités les plus longues sont aussi ceux qui obtiennent les diplômes les plus élevés : nous sommes ici face à un effet de structure ; une fois contrôlé le diplôme d'Ego, nous obtiendrons plus bas un tout autre résultat.

Au total, nous pouvons retenir de cette analyse préliminaire que la mise en couple en cours de scolarité ne paraît pas toujours correspondre à une rencontre liée aux études, et ne renforce pas toujours très fortement l'endogamie. Ainsi, si l'effet positif de la mise en couple en cours de scolarité sur l'endogamie est très net lorsque les deux conjoints sont scolarisés, il est d'ampleur plus faible lorsqu'un seul des conjoints est scolarisé à la mise en couple. Étant donné que le premier cas est bien moins fréquent que le second, on peut donc s'attendre à observer une influence limitée de ce facteur en recourant à des modèles

23 Chez les deux sexes, l'association est hautement significative selon un test du Khi2.

biographiques sur l'ensemble des cohortes de l'enquête Étude de l'histoire familiale : même si la proportion de mises en couple en cours de scolarité a nettement augmenté, nous n'avons pas de raisons de penser que les cas dans lesquels les deux conjoints sont scolarisés aient pris au sein de ces configurations une part plus importante²⁴.

Figure 6.5 : Relation entre les diplômes des conjoints selon la scolarisation à la mise en couple



Lecture : 80 % des premiers couples des femmes formés pendant les études des deux conjoints sont endogames

Source : enquête Biographies et entourage (Ined)

Champ : premiers couples des individus nés en France entre 1930 et 1950 et résidant en Île-de-France en 2001

4 Modéliser les chances de mise en couple : méthode et variables

Une analyse plus approfondie des hypothèses élaborées au début de ce chapitre nécessite un appareil statistique relativement élaboré. Nous commençons par présenter les

²⁴ Rappelons que l'enquête Étude de l'histoire familiale ne nous permet pas de connaître le statut du conjoint lors de la mise en couple : il nous est impossible de connaître l'évolution de la proportion de couples dans lesquels les deux conjoints étaient scolarisés lors de la mise en couple au fil des cohortes.

choix méthodologiques que nous avons réalisés, visant à répondre aux limites des techniques utilisées par les travaux précédents, avant d'en aborder les résultats.

4.1 Le modèle de Cox à risques concurrents

Nous utilisons un modèle à risques concurrents, et plus précisément un modèle de Cox à chances proportionnelles (*cf.* Encadré 6.1 ci-dessous) incluant des variables explicatives dépendantes du temps²⁵, qui consiste à modéliser les chances d'occurrence de chaque type d'événement possible (ici, la mise en couple avec un conjoint potentiel de chacun des niveaux de diplôme)²⁶, sachant qu'aucun autre événement (mise en couple) n'a encore eu lieu pour un individu donné. Nous retenons la stratégie d'estimation proposée par Mary Lunn et Don McNeil (1995 ; voir aussi Putter, Fiocco et Geskus, 2007), en introduisant simultanément dans un modèle unique l'ensemble des types d'événement possibles. Pour chaque individu, une entrée est créée concernant chaque type d'événement, les événements n'ayant pas eu lieu étant marqués comme censurés à la date d'occurrence de celui qui a eu lieu ; les individus n'ayant jamais vécu en couple à la date de l'enquête sont marqués comme censurés à cette date pour tous les types d'événements. Ces entrées donnent chacune lieu à une ligne par année observée avant l'événement ou la censure. Le format classique individu-année est donc ici remplacé par un format individu-type d'événement-année.

Le recours à un modèle à risques concurrents permet de ne pas séparer l'analyse des circonstances et du rythme de mise en couple (« quand ? ») de celle de l'homogamie, c'est-à-dire du résultat de cette mise en couple (« qui ? »). Une approche alternative (Smeenk, 1998, p. 46-47 ; de Graaf *et al.*, 2003, p. 97) consiste à analyser ces phénomènes en deux étapes séparées, d'abord à l'aide d'un modèle d'analyse des biographies, puis à l'aide d'une régression logistique multinomiale portant sur les couples finalement formés. Or, il nous semble essentiel de tenir compte du fait qu'un individu ne se met pas en couple de manière abstraite (Blossfeld et Timm, 2003c, p. 20) : il le fait avec une personne particulière, et il ne l'aurait probablement pas fait avec une autre, du moins pas dans les mêmes circonstances ; comme on l'a déjà défendu, la structure du marché conjugal sur lequel l'individu évolue

25 *Competing risks Cox proportional hazards model with time-dependent covariates.*

26 *Cause-specific hazards (CSH).* La principale approche alternative, celle par les *subdistribution hazards (SH, Fine et Gray, 1999)*, soulève des difficultés importantes lors de l'introduction de variables dépendantes du temps (Latouche, Porcher et Chevret, 2005 ; Beyersmann et Schumacher, 2008 ; Cortese et Andersen, 2010).

semble déterminante. En d'autres termes, en matière de mise en couple, « *l'occasion fait le larron* » (Bozon et Héran, 1988, p. 135). Et ce, d'autant plus qu'il est difficile de déterminer quand un individu recherche un conjoint et quand il ne le fait pas, tant cette recherche peut avoir comme support la plupart des activités sociales ordinaires, dont notamment les études, le travail ou encore les sorties (Oppenheimer, 1988). Les modèles à risques concurrents permettent de modéliser cette interaction entre le contexte de mise en couple et le choix du conjoint.

Néanmoins, tous les modèles à risques concurrents ne semblent pas également adaptés à notre analyse : le modèle retenu doit aussi permettre de contraster l'effet qu'a une variable sur un événement d'un certain type avec celui qu'elle a sur un autre, tout en offrant des indications concernant l'ampleur et la significativité de ces contrastes. La méthode retenue par Hans-Peter Blossfeld et Andreas Timm (2003a) ne possède aucune de ces deux propriétés, et mène donc à des observations peu instructives, comme le fait que la scolarisation diminue tout à la fois les chances de mise en couple endogame, hypergame et hypogame – alors la question est bien : sur quel type de couple la scolarisation a-t-elle l'effet négatif *le plus fort* ? Au contraire, la modélisation de Lunn et McNeil, si elle est strictement équivalente à l'estimation séparée d'un modèle pour chaque type d'événement possible, offre par rapport à cette dernière approche un avantage décisif : elle contraste automatiquement les chances de mise en couple d'un type donné avec un type de référence ; et surtout elle permet de calculer la significativité de ces contrastes – ce qui est impossible avec des modèles séparés fondés sur le même échantillon. En pratique, les effets différents selon le type de la mise en couple sont mesurés par le biais d'interactions entre les variables explicatives et une variable indiquant le type d'événement (que celui-ci ait réellement eu lieu, ou qu'il soit censuré, comme c'est le cas à la fois lorsque l'individu ne s'est jamais mis en couple, et lorsqu'il s'est mis en couple avec un conjoint d'un diplôme différent).

La définition des types d'événements concurrents constitue une autre caractéristique importante du modèle. À la suite de Jani Erola, Juho Härkönen et Jaap Dronkers (2012 ; Erola et Härkönen, 2013), nous avons choisi de retenir comme risques concurrents les niveaux *absolus* de diplôme des conjoints potentiels plutôt que leur niveau relatif (endogamie, hypergamie ou hypogamie). Si les deux approches peuvent se justifier, celle

retenue ici présente l'avantage d'une plus grande clarté d'interprétation²⁷. Par exemple, dans l'approche par niveau relatif, un renforcement de l'hypergamie doit être interprété comme une diminution de l'endogamie si on observe en même temps un renforcement de l'hypogamie, mais comme une augmentation générale du niveau de diplôme moyen des conjoints si on observe en parallèle un affaiblissement de l'hypogamie. Dans notre approche, ces deux questions sont clairement séparées grâce à une extension du modèle proposé par les auteurs déjà cités combinant deux dimensions du marché conjugal. La première dimension concerne la question de l'hypergamie et de l'hypogamie, autrement dit la dimension ordonnée du choix du conjoint, qui est fondamentalement asymétrique entre les deux conjoints, et par extension entre hommes et femmes. La seconde dimension est binaire plutôt que hiérarchique : elle oppose le choix d'un conjoint de diplôme identique au sien (endogamie), à tout autre choix (exogamie) ; cette dimension n'est pas intrinsèquement genrée du fait de l'égalité des diplômes de l'homme et de la femme au sein du couple.

Le modèle mettant en concurrence les niveaux de diplômes absolus des conjoints potentiels analyse par construction la première dimension (ordonnée). Afin de tenir en même temps compte de la seconde dimension (binaire), nous introduisons dans le modèle une variable dichotomique indiquant si l'événement considéré (censuré ou non) correspondrait à la formation d'un couple endogame s'il avait lieu. L'interaction entre cette variable et les autres variables explicatives permet d'analyser leur effet sur les chances de former un couple endogame. Prolongeant le principe au fondement de la méthode de Lunn et McNeil, cette solution revient en fait à estimer un modèle séparé pour les mises en couples endogames (qu'elles soient réellement observées ou censurées), alors que les cas d'exogamie sont divisés en un modèle pour chaque niveau de diplôme du conjoint potentiel.

Pour finir, le modèle de Cox permet de contrôler très finement la distribution temporelle des mises en couple pour chaque groupe de diplôme (*cf.* Encadré 6.1). Nous estimons ici des courbes de séjour distinctes (strates) pour chaque groupe de diplôme d'Ego, de manière à contrôler des différences de rythme de mise en couple liées au diplôme qui ne se résument

27 La difficulté à interpréter une modélisation en termes relatifs a été soulignée de longue date par Peter Blau et Otis Dudley Duncan (1967, chap. 4 et 5) à propos de l'analyse de la mobilité sociale intergénérationnelle. Ce point a été repris à propos de la mise en couple par de N. D. de Graaf *et al.* (2003, p. 97) et par Tak Wing Chan et Brendan Halpin (2003), soulignant en particulier le problème lié aux effets de plancher et de plafond (les individus non diplômés ne peuvent être hypogames, les plus diplômés ne peuvent être hypergames, mais ces deux groupes peuvent néanmoins choisir des conjoints plus ou moins diplômés).

pas à une multiplication ou une division des chances instantanées de mise en couple (non-respect de l'hypothèse de proportionnalité des chances). Ceci est essentiel dans l'optique de mesurer l'effet des circonstances biographiques sur le diplôme du conjoint une fois contrôlé le diplôme d'Ego. Une mauvaise prise en compte de l'effet de cette variable mènerait à des conclusions fortement biaisées puisque les plus diplômés – qui du fait de la tendance à l'homogamie sont aussi ceux qui prennent le plus souvent des conjoints eux-mêmes diplômés –, se mettent en couple plus tard, et plus souvent en cours de scolarisation.

Contrepartie de ces avantages, alors que nos données sont annuelles (*cf. supra*), le modèle de Cox est un modèle en temps continu : deux individus ne sont pas supposés pouvoir se mettre en couple exactement au même âge, ce qui arrive pourtant en pratique puisque nous ignorons le mois, le jour voire l'heure de la mise en couple. Le traitement de cette imperfection d'information nécessite de réaliser une approximation de la vraisemblance du modèle²⁸. Nous recourons ici à l'approximation d'Efron, considérée comme bien plus fiable que l'approximation de Breslow utilisée classiquement, et qui fournit des estimations précises même en présence d'une part importante d'événements simultanés (Therneau et Grambsch, 2000, p. 48).

Les modèles sont estimés dans l'environnement R (R Core Team, 2014), à l'aide du paquet *survival* (Therneau, 2014 ; Therneau et Grambsch, 2000), ainsi que du paquet *rms* (Harrell, 2015) pour la construction des bases des splines cubiques restreintes. Les intervalles de confiance, les erreurs-types et les probabilités critiques sont obtenues par l'estimateur « sandwich » d'Huber-White, et sont donc robustes à l'hétéroscédasticité²⁹.

Encadré 6.1 : Le modèle de Cox

Le modèle semi-paramétrique à chances proportionnelles de Sir David Cox (1972) est un outil classique de l'analyse des biographies. L'originalité de ce modèle porte sur la manière de modéliser la courbe de séjour de référence – courbe représentant à chaque période la proportion d'individus ayant connu l'événement étudié, n'en ayant pas encore fait l'expérience, ou dont l'observation a été censurée avant de l'avoir connu. Au lieu de faire l'hypothèse, souvent difficile à justifier théoriquement et empiriquement, que cette

28 Bien qu'une méthode exacte de calcul de la vraisemblance existe, son coût s'avère prohibitif même pour des échantillons de taille très modeste par rapport au nôtre.

29 Ces erreurs-types tiennent compte du fait qu'un seul événement est observé pour chaque individu.

courbe suit une distribution paramétrique, le modèle se contente de l'estimer de manière non paramétrique à partir des données observées.

Le modèle repose en revanche sur une hypothèse plus restrictive concernant l'effet des variables explicatives sur les chances instantanées de connaître l'événement, justifiant l'appellation « semi-paramétrique » du modèle. De manière similaire à l'hypothèse de linéarité des effets utilisée par les modèles de régression linéaire, le modèle de Cox permet au chercheur de résumer l'effet d'une variable par un seul coefficient, en faisant l'hypothèse qu'il agit de manière multiplicative sur les chances instantanées *de la même manière à toutes les périodes*. En d'autres termes, la courbe de séjour de référence est adaptée selon les caractéristiques de chaque individu, mais sa forme ne change pas : le rapport entre les chances instantanées de deux profils d'individus reste constant pour toutes les périodes. Selon cette hypothèse, les courbes de séjour de deux profils d'individus ne peuvent se croiser : les augmentations et diminutions des chances doivent être parallèles.

Cette spécification revient à confondre sous un même coefficient un effet d'accélération (l'événement arrive plus tôt) et un effet d'augmentation de la probabilité finale de l'événement étudié (l'événement arrive plus souvent). Sous l'hypothèse de proportionnalité des chances, ces deux phénomènes sont équivalents ; l'un entraîne l'autre. Puisque aucune interprétation n'est plus justifiée que l'autre, et pour éviter d'induire le lecteur en erreur, nous parlerons ici de manière générique d'« augmentation des chances de mise en couple ».

En cas de violation de cette hypothèse, on s'expose au risque d'interpréter une accélération comme une augmentation de la probabilité finale de l'événement (coefficient positif), un retard comme une diminution de cette probabilité (coefficient négatif), et vice-versa (Yamaguchi, 1992 ; Bernardi, 2001). Soulignons que ce danger n'est pas propre au modèle de Cox : il affecte aussi les modèles les plus couramment utilisés, dont les modèles exponentiels à chances constantes par morceaux, ainsi que les modèles en temps discret. Le modèle de Cox présente l'avantage non négligeable de fournir des outils permettant de tester l'hypothèse de proportionnalité, par le biais en particulier de l'étude des résidus de Schoenfeld, plutôt que d'en faire comme souvent un postulat non contrôlé. Dans les cas où l'hypothèse n'est pas respectée, il est possible d'introduire des

interactions entre les variables explicatives et le temps (ici, l'âge d'Ego), ce qui revient à autoriser le rapport entre les chances instantanées de deux profils d'individus à varier de manière complexe au cours du temps. C'est ce que nous faisons ici, de manière non seulement à respecter les hypothèses du modèle, mais surtout à analyser en détail les évolutions, non linéaires avec l'âge, de l'effet des variables d'intérêt.

Une seconde manière de respecter l'hypothèse de proportionnalité est d'introduire dans le modèle une courbe de séjour de référence pour chaque groupe défini par une variable explicative (strates). Puisqu'il est particulièrement important de contrôler correctement l'effet du niveau d'éducation, qui est la variable principale affectant le rythme de mise en couple, nous introduisons une strate pour chaque groupe de diplôme.

Pour finir, deux avantages supplémentaires du modèle de Cox sont qu'il modélise sans difficulté les chances d'occurrence d'un événement dont seule une partie de l'échantillon fera l'expérience (célibataires définitifs, *cf.* chapitre précédent), et qu'il prend en charge sans difficulté les variables dépendantes du temps, en introduisant une ligne pour chaque période observée (se reporter au corps du texte). Dans ce dernier cas, l'hypothèse de proportionnalité s'applique toujours, mais de manière plus complexe : les profils d'individus entre lesquels les chances doivent être proportionnelles deviennent des profils d'individus-année. On suppose alors par exemple que la scolarisation présente le même effet sur les chances de mise en couple quel que soit l'âge d'Ego, ce qui est validé par l'examen des données une fois les autres variables contrôlées.

4.2 Variables

4.2.1 La mesure du niveau d'éducation : causalité et anticipation

La question du contrôle de l'effet du diplôme nous amène à celle de la définition des variables explicatives. Il est courant dans les travaux utilisant des modèles d'analyse des biographies (Blossfeld, 1995 ; Blossfeld et Timm, 2003a) de contrôler le niveau d'éducation par une variable continue mesurant à chaque âge le nombre d'années passées dans le système scolaire. Outre que cette solution suppose des effets linéaires de l'éducation, qu'elle nie les différences entre filières pour un même niveau d'études (Breen et Jonsson, 2000 ; cette position est retenue par exemple par Breen *et al.*, 2009, p. 1477), et qu'elle repose le plus souvent sur des imputations d'une trajectoire « normale » du fait de l'absence de

données suffisantes, elle revient fondamentalement à ignorer que les individus ne sont pas myopes, c'est-à-dire qu'ils sont capables d'une certaine anticipation, notamment concernant le diplôme qu'ils obtiendront, à partir d'informations dont le chercheur ne dispose le plus souvent pas (Bocquier, 2006)³⁰. Il paraît naïf de considérer qu'un individu n'a aucune idée, en se mettant en couple avec un conjoint encore scolarisé, du niveau de diplôme que celui-ci atteindra, et que son style de vie ne porte aucun indice de ce statut futur – d'autant que le temps écoulé entre les deux événements est généralement court³¹. Puisque nous ne pouvons savoir quel est le niveau de connaissance et d'anticipation des individus, mais qu'il est essentiel pour notre analyse de contrôler aussi finement que possible les effets de la carrière scolaire d'Ego, nous introduisons séparément deux aspects de celle-ci dans nos modèles (une distinction retenue par Smeenk, 1998, p. 135, note 5) : d'une part, son résultat, défini de manière stable au cours du temps pour chaque individu ; d'autre part, son aspect biographique, variant chaque année observée.

Ainsi, comme expliqué plus haut, nous définissons les types d'événement – soit le niveau d'éducation du conjoint, ainsi que la nature endogame ou exogame de l'union – à partir des diplômes finalement obtenus par Ego et son premier conjoint. Cette définition nous permet d'étudier les mécanismes produisant *in fine* l'homogamie de diplôme, plutôt que de s'en tenir à un instantané pris au moment de la mise en couple. Son anachronisme apparent ne crée pas de difficulté particulière, puisque les variables biographiques d'intérêt restent, elles, définies par rapport au statut d'Ego une année donnée. Le diplôme d'Ego et

30 Plus généralement, la position épistémologique retenue par H.-P. Blossfeld et A. Timm (2003a, p. 13-14 et 20-21 ; voir aussi Blossfeld et Rohwer, [1995] 2002, chap. 1 et 6 ; Pötter et Blossfeld, 2001) concernant l'identification de la causalité à partir de données biographiques nous paraît naïve du point de vue de sa mise en œuvre pratique. Un événement ayant lieu au temps t' peut *statistiquement* présenter un « effet » sur un événement ayant eu lieu en t (avec $t < t'$), pour la raison tout à fait classique qu'un biais lié à l'existence d'un troisième événement non contrôlé (*confounding variable*), encore antérieur, et corrélé aux deux autres, peut toujours exister. Il est ainsi banal d'observer que les individus qui ont pris (en t) pour conjoints des personnes diplômées obtiendront eux-mêmes finalement (en t') un diplôme élevé. Il n'est pas nécessaire de supposer que le conjoint a aidé, par son niveau scolaire élevé, à l'obtention du diplôme d'Ego, pour expliquer ce phénomène ; mais seulement de noter que les Ego en question ont aussi une origine sociale plus élevée que la moyenne, « hétérogénéité non observée » qu'il est très difficile de contrôler complètement, mais dont l'effet, conscient ou non pour les acteurs concernés, est très fort. Le recours à des approches longitudinales ne permet donc pas ici de résoudre complètement les problèmes de corrélation fallacieuse dont souffrent les méthodes transversales (Courgeau et Lelièvre, 1989, p. 72-75 ; Yamaguchi, 1991, chap. 6).

31 Cette question de l'incertitude concernant les caractéristiques futures du conjoint (mais aussi d'Ego par rapport à ses propres caractéristiques et préférences futures) est au centre de l'approche développée par Valerie Oppenheimer (1988, p. 565), dans laquelle elle est supposée affecter fortement le rythme de mise en couple.

celui de son conjoint sont mesurés en quatre catégories³² : aucun diplôme ou certificat d'études primaires (CEP) ; certificat d'aptitude professionnelle (CAP), brevet d'enseignement professionnel (BEP), brevet d'études du premier cycle (BEPC) ou assimilé ; baccalauréat général, technique ou professionnel ; diplôme du supérieur. La variable indiquant si la formation d'un couple avec un conjoint d'un diplôme donné serait endogame ou non (voir ci-dessus) est créée à partir de ces deux variables de diplôme. Le diplôme finalement obtenu par Ego est aussi utilisé comme variable explicative afin de contrôler le rythme de mise en couple, ainsi que la tendance à choisir chaque type de conjoint, des différents groupes de diplôme³³.

4.2.2 Origine sociale

L'origine sociale est introduite à titre de contrôle, en plus du diplôme d'Ego. On peut en effet supposer que les individus d'origine sociale élevée se mettront en couple avec des individus plus diplômés du fait d'une hétérogénéité à l'intérieur des niveaux de diplôme, mais aussi grâce à des capitaux économique, culturel et social plus élevés. Ceci pourrait affecter les résultats si ces individus ont aussi un comportement différent des autres du point de vue de la mise en couple (Goldscheider et Waite, 1986 ; Cooney et Hogan, 1991 ; Axinn et Thornton, 1992) – par exemple du fait d'une moindre difficulté à acquérir un logement pour se mettre en ménage, ou à des normes sociales différentes selon les groupes sociaux. Nous contrôlons ainsi la dernière profession occupée par le père d'Ego (le cas échéant, avant la retraite), ou, à défaut, par sa mère (environ 2 % des individus), en six catégories socioprofessionnelles, plus une catégorie comprenant les personnes n'ayant jamais travaillé ainsi que les valeurs manquantes, de manière à ne pas biaiser l'échantillon en omettant trop d'individus n'ayant pas répondu à cette question.

32 Le choix d'une classification très agrégée est dicté par la faible significativité de différences trop fines entre de nombreux groupes, ainsi que par la volonté de simplifier la présentation de résultats déjà complexes à lire. Néanmoins nous obtenons des résultats similaires à l'aide de classifications plus détaillées.

33 L'intérêt de cette variable est justifié empiriquement par la très nette baisse des indicateurs BIC et AIC lors de l'introduction du diplôme, par rapport à un modèle mesurant le niveau scolaire à travers le seul temps passé dans le système éducatif (variable maintenue égale à la durée des études d'Ego une fois celui-ci sorti du système éducatif).

4.2.3 Variables biographiques d'intérêt : temps et scolarisation

Contrairement aux variables de contrôle stables au cours de la vie d'Ego que nous venons de décrire, nos variables d'intérêt sont proprement biographiques en ce qu'elles décrivent l'état d'Ego une année donnée, sans introduire d'informations concernant son état futur.

La scolarité d'Ego est mesurée, selon les modèles, de deux manières. Premièrement, par le statut d'Ego, défini en trois modalités : scolarisé ; non scolarisé et n'ayant jamais travaillé ; ayant déjà travaillé pendant plus de trois mois³⁴. Le second statut, regroupant les individus n'étant pas stabilisés dans la vie active, vise à repérer éventuellement un effet propre de l'activité professionnelle, qui ne se définirait ainsi pas uniquement comme le négatif de la scolarité ; le faible nombre de mises en couple observées sous cette modalité rend cependant peu probante la mesure de ses effets. Deuxièmement, l'effet de la scolarisation est appréhendé par le temps passé par Ego dans le système scolaire depuis ses 14 ans. Cette variable est égale à 1 à 15 ans, et elle est incrémentée d'un an chaque année, pour finalement prendre la valeur zéro après la fin des études d'Ego³⁵. Elle vise à approcher la progression dans le système scolaire³⁶, et en particulier le fait que les fréquentations d'Ego peuvent être supposées devenir de plus en plus homogènes du point de vue du niveau de diplôme final à mesure que se renforce la sélection scolaire, et redevenir plus hétérogènes après la fin des études. La comparaison des effets de ces deux variables nous permettra de tenter de préciser le mécanisme en jeu.

Deux variables explicatives supplémentaires sont utilisées alternativement dans les modèles, et concernent l'évolution au cours du temps des chances de mise en couple. La première est l'âge d'Ego, dont on a souligné qu'il recouvre indissociablement une situation biographique, et notamment le temps écoulé depuis la fin de la scolarité et le premier emploi³⁷, et une position sur le marché conjugal influençant les opportunités de rencontre.

34 L'enquête Étude de l'histoire familiale ne contient pas d'informations plus précises concernant le statut d'emploi une année donnée.

35 Cette remise à zéro est nécessaire puisque nous ne cherchons pas à mesurer le niveau d'éducation d'Ego, mais sa fréquentation d'un niveau scolaire donné.

36 Rappelons que nous ne disposons pas d'informations concernant la carrière scolaire exacte d'Ego, mais seulement concernant son âge de fin d'études.

37 En effet, la soustraction à l'âge du temps passé en scolarisation suffit à déterminer complètement le temps écoulé depuis la fin de la scolarité. De plus, la corrélation entre le temps écoulé depuis l'entrée en activité et l'âge, supérieure à 0,8, ne permet pas d'introduire simultanément ces deux variables dans un modèle contrôlant de plus le diplôme d'Ego. Certaines questions théoriques concernant l'effet du temps écoulé

Afin de pouvoir observer des évolutions non linéaires, nous utilisons des splines cubiques restreintes à deux degrés de liberté (*cf.* Encadré 5.2, p. 308), les trois nœuds (19, 23 et 28 ans) étant choisis pour correspondre approximativement aux quantiles 10 %, 50 % et 90 % de l'âge de mise en couple (Harrell, 2001, p. 23), tout en étant identiques chez les deux sexes de manière à faciliter les comparaisons. La seconde variable, le temps écoulé depuis la fin des études, est égale à zéro pendant la scolarité d'Ego, et augmente d'une unité chaque année après sa sortie du système scolaire. Elle vise à mesurer une potentielle augmentation de l'hétérogénéité des réseaux sociaux avec l'éloignement du contexte scolaire.

4.2.4 Année de naissance

Enfin, nous mettons l'ensemble des variables explicatives en interaction avec l'année de naissance pour modéliser les évolutions au fil des cohortes. Puisque, au vu des évolutions de la mise en couple déjà évoquées au chapitre précédent, nous n'avons aucune raison de penser que l'effet des variables ait été linéaire, nous avons ici encore recours à des splines cubiques restreintes à deux degrés de liberté (nœuds placés en 1922, 1945 et 1968). Une spécification linéaire risquerait en effet de faire apparaître des artefacts sur certaines cohortes du fait d'un mauvais contrôle de l'évolution curvilinéaire de l'effet de certaines variables.

4.3 Spécification des modèles

Trois modèles différents sont successivement considérés afin de répondre aux questions développées plus haut. Le modèle 0 comprend l'ensemble des variables indépendantes du temps, mais aucune des variables biographiques d'intérêt. Il est utilisé comme référence, pour évaluer l'apport réel de la dimension biographique à la compréhension de la mise en couple. Partant de cette spécification de base, le modèle 1 cherche à mettre en évidence un effet de la scolarisation et de l'âge sur le choix du conjoint, au-delà des variables de contrôle présentées ci-dessus. Finalement, le modèle 2 vise à préciser les résultats du précédent, en décomposant les effets de la scolarisation et de l'écoulement du temps. La première est mesurée à la fois comme dans le modèle 1 par le statut (scolarisé, n'a jamais travaillé, a déjà occupé un emploi plus de 3 mois), mais aussi par le temps de scolarisation. Le second est

depuis la fin des études ou le temps passé en emploi (indice de stabilité professionnelle) sont donc en pratique difficilement dissociables de l'âge, du point de vue des individus aussi bien que du chercheur.

mesuré par cette dernière variable et par le temps écoulé depuis la fin des études. Si les modèles 1 et 2 sont emboîtés dans le modèle 0, le modèle 2 ne l'est pas dans le modèle 1, en dépit de sa plus grande complexité : l'effet du temps est en effet spécifié différemment dans les deux modèles, aucune des deux versions ne pouvant se ramener à l'autre.

Les formules prédisant les chances instantanées de mise en couple avec un conjoint d'un diplôme donné sont donc, dans la notation initiée par Wilkinson et Rodgers (1973) :

$$M0 : \text{Chances} = \text{Année de naissance} \times \text{Endogamie} \times (\text{Diplôme final} + \text{Origine sociale})$$

$$M1 : \text{Chances} = \text{Année de naissance} \times \text{Endogamie} \times$$

$$(\text{Diplôme final} + \text{Origine sociale} + \text{Statut} + \text{Âge})$$

$$M2 : \text{Chances} = \text{Année de naissance} \times \text{Endogamie} \times$$

$$(\text{Diplôme final} + \text{Origine sociale} + \text{Statut}$$

$$+ \text{Années de scolarisation} + \text{Années depuis la fin des études})$$

Ces formules doivent être complétées en rappelant que des splines sont retenues pour modéliser les effets de l'âge et de l'année de naissance, et que des courbes de séjour différentes sont introduites pour chaque groupe de diplôme (strates)³⁸.

Ces modèles n'étant pas interprétables à l'aide de tableaux de régression classiques du fait du très grand nombre de coefficients, d'interactions complexes et de splines, nous représentons en appendice à ce chapitre les évolutions au fil des cohortes par des graphiques. Néanmoins, les mêmes modèles sans variation selon l'année de naissance ont été estimés sur les cohortes 1925-1944 et 1945-1964 : leurs résultats, qui confirment ceux obtenus à partir du modèle d'ensemble, sont fournis dans les tableaux 6.4 et 6.5 (p. 387 et 391). Dans ce qui suit, nous présenterons à l'aide de graphiques les résultats significatifs du modèle d'ensemble, mais renverrons le lecteur à ces tableaux pour ceux qui ne le sont pas.

38 De nombreuses variantes de ces modèles (une quarantaine au total) ont été estimées pour s'assurer de la robustesse des résultats présentés ici. En particulier, ceux-ci sont cohérents avec ce que l'on observe en introduisant des splines à trois degrés de liberté (plutôt que seulement deux) ou en retenant une spécification linéaire (un seul degré de liberté).

5 Résultats des modèles à risques concurrents

5.1 Validation des modèles

Nous recourons à plusieurs indicateurs adaptés aux modèles de durée (Harrell, 2001, p. 492) pour évaluer l'apport des modèles 1 et 2 par rapport au modèle 0.

Tout d'abord, le coefficient de détermination R^2 , dans la variante de Nagelkerke. Celle-ci est fondée sur le R^2 généralisé de Cox et Snell, qui compare la vraisemblance du modèle considéré à celle du modèle vide (n'incluant aucune variable explicative), soit : $R_{CS}^2 = 1 - \left(\frac{L_0}{L}\right)^{2/n}$ avec L_0 la vraisemblance du modèle vide, L la vraisemblance du modèle considéré, et n le nombre d'événements observés. Cette quantité ayant un maximum inférieur à 1, la variante de Nagelkerke consiste à rapporter cette quantité à sa valeur maximale $1 - L_0^{2/n}$. Dans le cas présent, où les modèles comprennent une strate pour chaque niveau de diplôme, le modèle de référence n'est pas réellement vide : il tient déjà compte du diplôme des individus. Le R^2 est donc un indicateur assez exigeant, puisqu'il mesure l'apport des variables explicatives au-delà de l'effet du diplôme. Quoi qu'il en soit, c'est principalement la différence entre nos trois modèles qui retiendra notre attention, plus que la valeur absolue du R^2 .

Le second indicateur utilisé est l'indice de concordance D de Somers (Liebetrau, 1983, p. 76). Celui-ci est fondé sur la construction de l'ensemble des paires de deux individus qui peuvent être formées à partir de l'échantillon ; il est égal à la différence entre la proportion de paires dans lesquelles l'ordre des âges de mise en couple prédits par le modèle est en accord avec celui des âges réellement observés (ordre au sens de $t_1 < t_2$ pour deux individus 1 et 2) et celle des paires dans lesquelles l'ordre est en désaccord. Ne sont donc pas prises en compte les paires dans lesquelles la mise en couple ou la censure ont lieu en même temps, de même que les paires dans lesquelles aucun des deux individus ne s'est mis en couple à la date de l'enquête. En revanche, lorsque la mise en couple a été observée pour l'un des individus de la paire *avant* que la censure n'intervienne pour l'autre individu, le caractère concordant ou discordant de la paire peut être observé. La référence est ici un modèle réellement vide : $D = 0$ si le modèle n'a aucune capacité prédictive (prédiction au hasard) ; à l'inverse, $D = 1$ pour un modèle prédisant parfaitement l'âge de mise en couple.

Enfin, les indicateurs AIC et BIC, déjà utilisés dans les chapitres précédents, sont eux aussi utilisés pour estimer l'apport des modèles successifs. Le nombre d'événements est utilisé pour calculer le BIC, et non le nombre d'épisodes qui est ici extrêmement élevé par rapport aux mises en couple réellement observées (Raftery, 1995).

La comparaison des modèles 1 et 2 au modèle 0 montre que l'apport des variables biographiques est modeste mais bien réel (tableau 6.3). Le R^2 et l'indice de concordance indiquent que le modèle 1, malgré sa complexité moindre, explique mieux la mise en couple que le modèle 2 ; logiquement, les indicateurs AIC et BIC vont dans le même sens. De fait, on montrera ci-dessous que la prise en compte du temps de scolarisation, qui constitue l'un des deux apports du second modèle n'apporte pas de contribution très claire à l'explication de la mise en couple ; seul le temps écoulé depuis la fin des études présente un effet significatif. Nous présentons néanmoins les résultats des deux modèles : l'effet du temps écoulé depuis la fin des études semble plus aisément interprétable dans un modèle contrôlant d'abord le temps passé en scolarisation (modèle 2), même si son introduction dans le modèle 1 représente aussi une amélioration selon les indicateurs AIC et BIC.

L'hypothèse de proportionnalité des chances de mise en couple qui est au fondement du modèle de Cox (Encadré 6.1) apparaît largement respectée. Si, du fait de la grande taille de notre échantillon, les corrélations entre l'âge d'Ego et les résidus de Schoenfeld sont parfois significatives, les coefficients de corrélation sont toujours inférieurs à 0,05 (et, pour l'immense majorité, à 0,02), signe que les violations de la proportionnalité sont négligeables. Ce bon ajustement du modèle peut probablement être attribué à l'utilisation de strates par niveau de diplôme, ainsi qu'à l'introduction d'interactions entre les variables temporelles (âge d'Ego, temps de scolarisation, années depuis la fin des études) et les différentes variables d'intérêt.

Tableau 6.3 : Statistiques d'ajustement des trois modèles

	Hommes			Femmes		
	Modèle 0	Modèle 1	Modèle 2	Modèle 0	Modèle 1	Modèle 2
Périodes	3 180 240	3 180 240	3 180 240	3 533 452	3 533 452	3 533 452
Événements	59 621	59 621	59 621	80 413	80 413	80 413
D.L.	149	182	194	149	182	194
Concordance	74,94 %	75,89 %	75,42 %	76,11 %	76,44 %	76,12 %
R ² de Nagelkerke	4,44 %	4,78 %	4,61 %	4,66 %	4,86 %	4,72 %
Log-vraisemblance	-541 657	-540 057	-540 862	-729 590	-728 363	-729 212
Log-vraisemblance du modèle vide	-562 857	-562 857	-562 857	-758 568	-758 568	-758 568
AIC	1 083 612	1 080 479	1 082 111	1 459 478	1 457 089	1 458 813
BIC	1 084 952	1 082 116	1 083 856	1 460 863	1 458 781	1 460 616

D.L. : Degrés de liberté

5.2 L'évolution au cours du temps du niveau de diplôme des conjoints : une convergence entre hommes et femmes

5.2.1 L'hypergamie augmente avec l'âge

L'âge, en tant qu'indicateur à la fois du statut d'Ego et de sa situation sur le marché conjugal, offre une première manière d'approcher le contexte de mise en couple. De fait, celui-ci présente des effets assez forts sur la rencontre d'un conjoint, et révèle des différences importantes entre les sexes et les cohortes (Figure 6.6). Commentons tout d'abord la situation des cohortes nées avant-guerre, avant de présenter les évolutions – majeures – qui ont pris place jusqu'aux cohortes nées dans les années 1960. Pour chaque période, nous interprétons les résultats à la lumière de l'évolution avec l'âge de la population des célibataires, de manière à mettre à l'épreuve à la fois les tendances prédites plus haut, et le lien entre les contraintes structurelles et l'évolution avec l'âge du choix du conjoint.

Parmi les cohortes nées dans les années 1930, une fois contrôlés le diplôme et l'origine sociale, un homme présente 6 fois plus de chances à 30 ans qu'à 20 ans³⁹ de se mettre en couple avec une diplômée du supérieur, par rapport à celles de rencontrer une femme non

39 Soit 30/5 : 30 fois plus de chances à 30 ans qu'à 15 ans, et 5 fois plus à 20 ans qu'à 15 ans.

diplômée ; ces chances continuent à augmenter légèrement au-delà de cet âge. Seule une évolution plus modeste est visible chez les femmes de ces générations : le facteur correspondant est de 1,6 à 30 ans⁴⁰, et il décroît de manière sensible ensuite. Le schéma est le même concernant les autres niveaux de diplôme des conjoints, bien que l'augmentation avec l'âge soit plus faible chez les deux sexes. Ces résultats confirment ceux obtenus à partir de données néerlandaises par W. Smeenk (1998, sect. 7.4.2.2), qui notait une augmentation des chances d'hypergamie et une diminution de celles d'hypogamie avec l'âge chez les hommes, alors que chez les femmes seules les chances d'hypogamie diminuaient par rapport à celles d'endogamie.

Ces tendances trouvent un reflet relativement fidèle dans l'évolution avec l'âge de la structure par diplôme de la population des célibataires du sexe opposé (Figure 6.7). Chez les hommes, on note dans un premier temps une légère hausse du niveau moyen de diplôme des célibataires, puis une baisse plus nette. Cette évolution correspond bien à la combinaison de deux effets : celui de l'âge de mise en couple, qui augmente – mais pas très nettement – avec le niveau de diplôme, pour la première tendance ; celui du taux de célibataires définitifs, qui baisse avec le diplôme, pour la seconde tendance. Le tableau est bien différent chez les femmes, chez qui l'on observe une élévation continue du diplôme moyen des célibataires, résultat de l'augmentation à la fois de l'âge de mise en couple et du taux de célibataires définitifs avec le niveau de diplôme. Les prédictions proposées plus haut sont donc entièrement validées.

Le lien entre ces évolutions structurelles et les chances de mise en couple est lui aussi confirmé. La forte augmentation, du côté des hommes, des chances de rencontrer une conjointe très diplômée, reflète clairement la hausse avec l'âge du diplôme moyen des femmes célibataires. Du côté des femmes, l'augmentation moins importante de ces chances, suivie d'une légère baisse, correspond elle aussi à l'évolution, peu marquée et en deux temps, que l'on observe chez les hommes.

Cette confirmation nous permet de répondre à nos interrogations concernant le fonctionnement du marché conjugal. Le modèle classique de la mise en couple, selon lequel les hommes devraient préférer des femmes moins diplômées et plus jeunes qu'eux lorsque leur position sur le marché leur donne cette latitude, n'est pas validé. Il semble bien que la

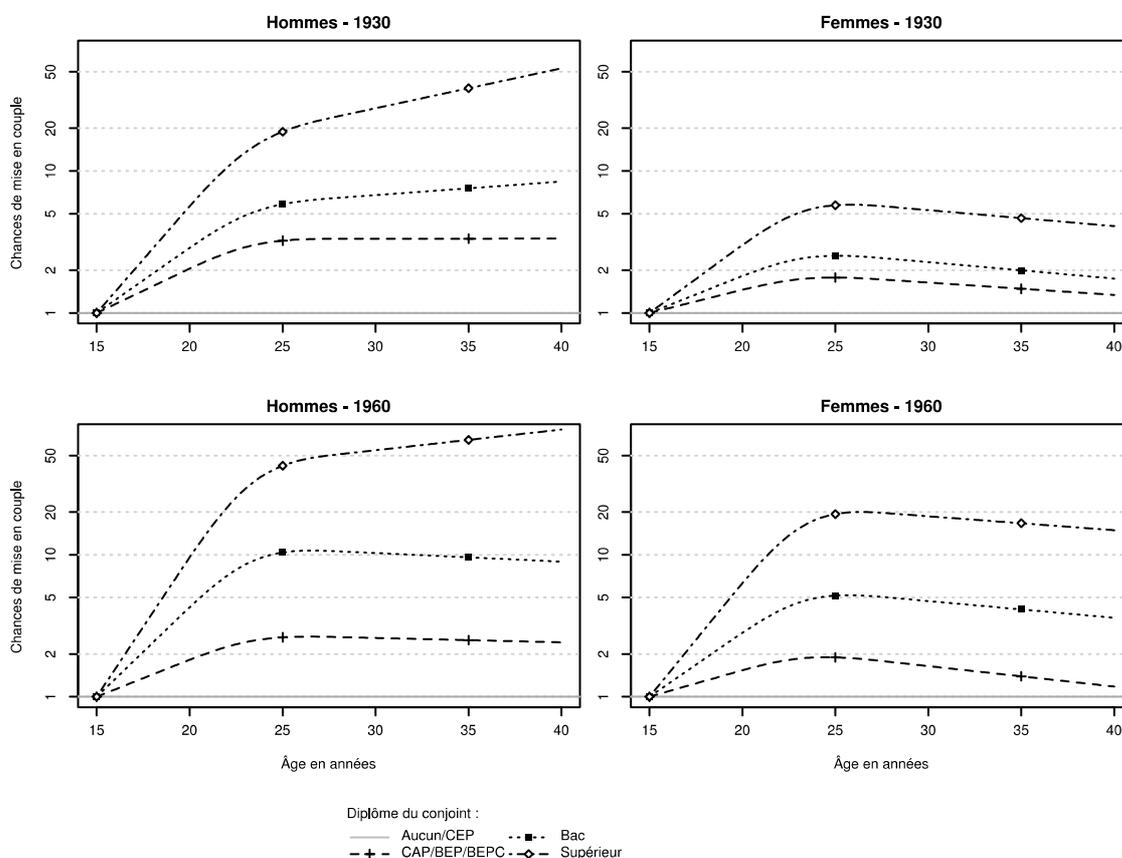
40 Soit 5/3 : 5 fois plus de chances à 30 ans qu'à 15 ans, et 3 fois plus à 20 ans qu'à 15 ans.

sur-représentation croissante des plus diplômées parmi les célibataires d'un âge donné soit plus déterminante pour le choix du conjoint des hommes qu'une norme d'hypergamie féminine dont on peine ici encore à trouver une preuve solide. Dans l'autre sens, interpréter l'augmentation avec l'âge des hommes du niveau moyen de diplôme de leurs conjointes comme le résultat d'une préférence pour la maximisation du statut du conjoint combinée à une position de force sur le marché conjugal semble exagéré : moins que d'un choix pour des conjointes très diplômées, ce phénomène semble surtout résulter d'une préférence pour des conjointes d'un âge relativement proche. C'est l'évolution de la structure de la population des célibataires des cohortes proches qui semble l'explication principale.

L'évolution avec l'âge observée chez les femmes amène quant à elle à relativiser l'idée selon laquelle celles-ci verraient leur valeur sur le marché conjugal se dégrader avec le temps. En effet, il semble important de souligner que les chances de mise en couple des femmes avec un conjoint très diplômé sont loin de *baisser* avec l'âge : tout au plus augmentent-elles moins vite que celles des hommes. La baisse que l'on observe après 25 ans est assez faible pour que ces chances soient à 40 ans encore supérieures à ce qu'elles étaient à 20 ans. Ce résultat est d'autant plus remarquable que la proportion de non diplômés parmi les hommes célibataires augmente assez nettement au-delà de 25 ans, ce qui ne se retrouve guère dans l'évolution des chances de mise en couple des femmes. Lorsque le choix du conjoint s'écarte des contraintes liées à la structure de la population des célibataires, il va à l'encontre de l'hypothèse d'une diminution de la valeur des femmes sur le marché conjugal.

Chez les deux sexes, on voit donc que le modèle fortement genré doit être mis en question même concernant les générations les plus anciennes : si on trouve bien des signes des mécanismes asymétriques évoqués plus haut, ceux-ci sont moins marqués qu'on aurait pu le penser, et la préférence pour l'hypergamie féminine n'est pas manifeste.

Figure 6.6 : Évolution avec l'âge des chances de rencontrer un conjoint d'un niveau de diplôme donné



Référence : 1 pour un conjoint non diplômé à 15 ans

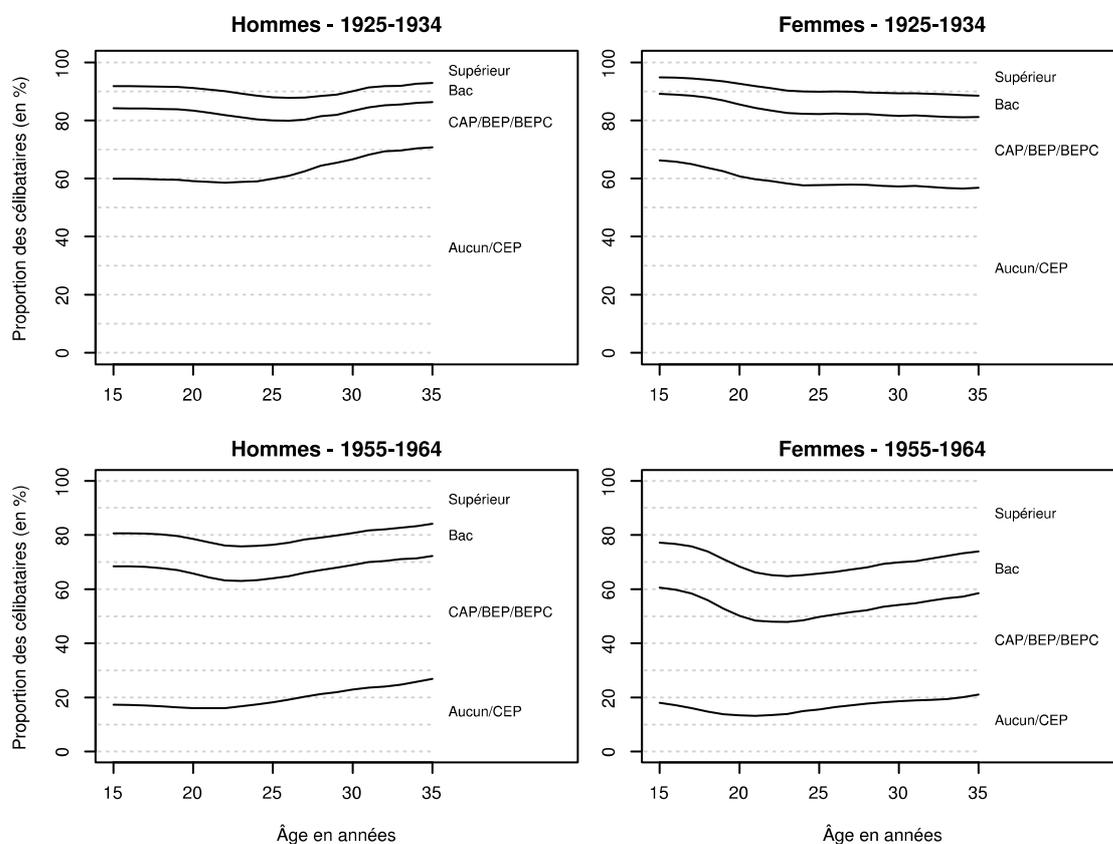
Le modèle 1 utilisé ici contrôle l'effet du diplôme final, de l'origine sociale et du statut au sein de chaque cohorte de naissance.

Lecture : Une femme née en 1960 a 5 fois plus de chances de rencontrer un diplômé du supérieur plutôt qu'un non diplômé à 20 ans qu'à 15 ans ; et 20 fois plus de chances à 30 ans qu'à 15 ans ; soit 4 fois (20/5) plus de chances à 30 ans qu'à 20 ans.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France et n'ayant pas connu de séparation

Figure 6.7 : Évolution avec l'âge de la structure par diplôme finalement obtenu de la population des célibataires (proportions cumulées)



Lecture : Parmi les femmes nées entre 1925 et 1934 et n'ayant jamais vécu en couple à 30 ans, 57 % n'obtiendront aucun diplôme ou le seul certificat d'études primaires, et 82 % obtiendront au plus un CAP, un BEP ou un BEPC.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : individus nés en France

La situation que nous venons de décrire, valable pour les cohortes nées avant guerre, s'est profondément modifiée jusqu'aux cohortes nées dans les années 1960 et 1970 : le modèle décrit plus haut, qui est celui retenu par de nombreuses analyses du marché conjugal, n'a plus cours. Chez les hommes nés autour de 1960, l'effet de l'âge sur les chances de rencontrer une conjointe diplômée du supérieur ou du baccalauréat s'est renforcé avant 30 ans, mais la progression ralentit ensuite nettement, voire se stabilise (Figure 6.6). Chez les femmes, l'évolution est frappante : d'une relativement faible évolution avec l'âge, on est passé à un schéma assez proche de celui valable chez les hommes des mêmes générations, quoique toujours avec une hausse plus faible et une décroissance plus marquée. Si les chances des hommes de rencontrer à 30 ans une diplômée du supérieur plutôt qu'une non diplômée par rapport aux chances à 20 ans ont légèrement baissé de la cohorte 1930 à la cohorte 1960, passant de 6 à 5 fois⁴¹, on passe chez les femmes de 1,6 à 3 fois⁴².

Ici encore, ces évolutions correspondent en partie à une modification avec l'âge de la composition par diplôme de la population des célibataires (Figure 6.7). Du côté des hommes, comme on l'a vu au chapitre précédent, l'âge de mise en couple des non diplômés s'est progressivement abaissé, pour finalement passer sous celui des diplômés du secondaire chez les cohortes nées après 1960 : on note donc une augmentation plus marquée avec l'âge du niveau moyen de diplôme des hommes célibataires. Cette augmentation s'inverse cependant assez nettement au-delà de 23 ans du fait de la hausse du taux de célibat définitif des hommes peu diplômés. On observe donc dans un premier temps une hausse plus importante qu'avant des chances d'hypergamie des femmes, puis une baisse après cet âge. Du côté des femmes, l'augmentation avec l'âge du niveau moyen de diplôme des célibataires est devenue plus marquée que pour les cohortes nées avant guerre⁴³, mais on note désormais une baisse progressive après 23 ans, liée à la disparition du surcroît de célibat des diplômées. L'augmentation des chances d'hypergamie des hommes est donc plus marquée avant cet âge, mais la stabilisation est plus nette au-delà.

Dans l'ensemble, la convergence relative entre hommes et femmes est donc due à une progression plus régulière de l'âge de mise en couple suivant le diplôme, et à une réduction des différences de taux de célibat selon le diplôme, entre les deux sexes. L'hypothèse d'une

41 Soit 50/10 : 50 fois plus de chances à 30 ans qu'à 15 ans, et 10 fois plus à 20 ans qu'à 15 ans.

42 Soit 18/6 : 18 fois plus de chances à 30 ans qu'à 15 ans, et 6 fois plus à 20 ans qu'à 15 ans.

43 Cette évolution semble due à la plus grande concentration de la mise en couple des plus diplômées (*cf.* chapitre précédent).

influence déterminante des contraintes structurelles sur l'évolution avec l'âge du choix du conjoint est donc validée. À l'inverse, les différences genrées apparaissent plus limitées qu'on n'aurait pu le penser au regard des théories classiques du couple, même au sein des cohortes nées avant la Seconde Guerre mondiale. Finalement, la norme d'hypergamie féminine ne paraît pas jouer un rôle observable sur les choix des individus, en particulier ceux des hommes âgés, supposés être en position de force sur le marché conjugal.

5.2.2 L'âge de mise en couple n'affecte pas significativement l'endogamie

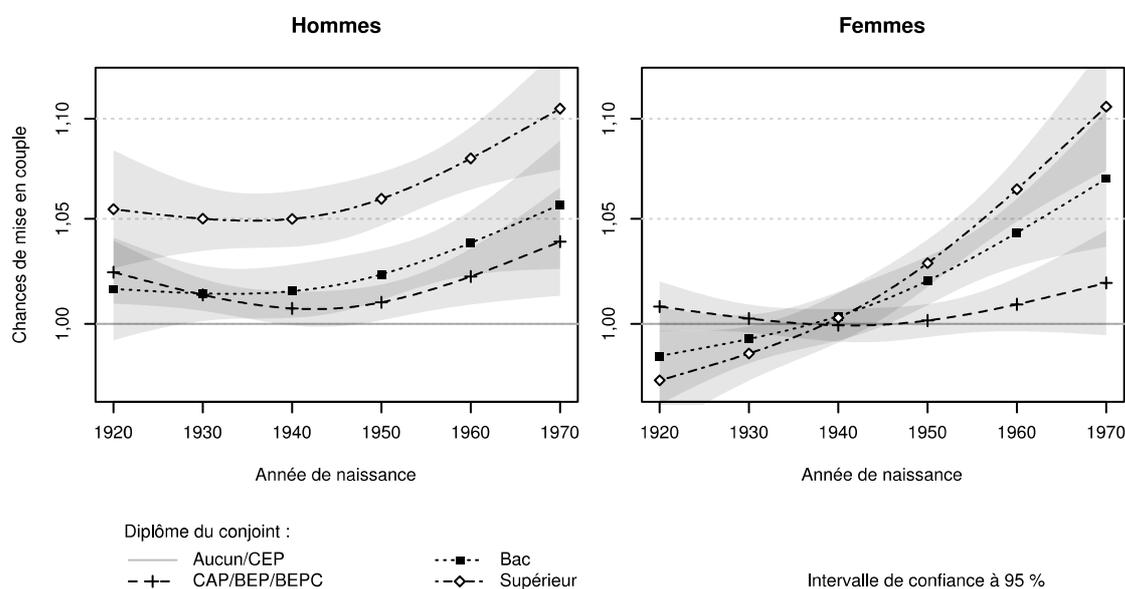
L'âge ne présente pas d'effet significatif sur les chances de mise en couple endogame (*cf.* tableau 6.4, p. 387). Néanmoins, les intervalles de confiance ne nous permettent pas d'exclure l'existence d'un effet qui pourrait, pour un écart d'âge d'une dizaine d'années, aller jusqu'à une augmentation ou une diminution de 20 % des chances de mise en couple endogame. Quoiqu'il en soit, il est clair que c'est le temps écoulé depuis la scolarisation qui a le plus fort effet sur l'endogamie, comme nous le verrons plus bas.

5.2.3 Années depuis la fin des études : une confirmation de l'effet de l'âge

Considérant que la fin des études marque une étape importante concernant le style de vie, la composition des réseaux sociaux et le passage à l'âge adulte d'un individu, il semble intéressant de confronter les résultats concernant l'âge que nous venons de présenter à ceux de notre second modèle, mesurant l'effet d'une année de plus après la sortie du système scolaire (Figure 6.8). Nos interprétations précédentes se trouvent pleinement confirmées.

Chez les hommes, le temps écoulé depuis la fin des études augmente les chances de rencontrer un conjoint très diplômé, et donc de faire un choix hypergame. Cet effet a gagné en importance au fil des cohortes : pour un homme né en 1920, les chances de se mettre en couple avec une diplômée du supérieur sont multipliées par 1,7 dix ans après la fin des études, par rapport à l'année de fin d'études ; pour un homme né en 1970, le facteur correspondant est de 2,6. On retrouve donc l'augmentation avec l'âge notée plus haut, mais avec une évolution inversée au fil des cohortes, qui semble indiquer que la fin des études a gagné en importance comme point de référence dans la trajectoire biographique des individus.

Figure 6.8 : Effet d'une année de plus après la fin des études sur les chances de rencontrer un conjoint d'un niveau de diplôme donné



Référence : 1 pour un conjoint non diplômé l'année de fin d'études

Le modèle 2 utilisé ici contrôle l'effet du diplôme final, de l'origine sociale, du statut et du temps de scolarisation au sein de chaque cohorte de naissance.

Lecture : Pour un homme ou une femme nés en 1970, à diplôme final contrôlé, chaque année supplémentaire après la fin des études augmente de 10 % les chances de rencontrer un conjoint diplômé du supérieur plutôt qu'un conjoint non diplômé ; soit des chances multipliées par 2,6 ($1,1^{10}$) dix ans après la fin des études.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France et n'ayant pas connu de séparation

Chez les femmes, l'évolution est plus nette encore : pour la cohorte 1920, les chances de rencontrer un diplômé du supérieur sont inférieures de 25 % dix ans après la fin des études, alors que pour la cohorte 1970 elles sont multipliées par 2,6. La baisse des chances ne s'observe que pour les cohortes de femmes nées avant 1940, et l'effet n'est significatif au seuil de 5 % que pour les cohortes 1920 à 1930 ; en outre, dans le modèle concernant les cohortes 1925-1944 (Tableau 6.5, p. 391) aucun effet n'est mesuré. Pour ces raisons ainsi que du fait des possibles biais liés à la mortalité différentielle déjà soulignés, il nous est difficile d'affirmer que l'on assistait parmi ces cohortes à une véritable décroissance plutôt qu'à une stabilité des chances de vie en couple avec un conjoint très diplômé – résultat qui viendrait nuancer la critique du modèle fortement genré formulée dans la section précédente. Le

schéma d'ensemble est dans tous les cas confirmé, que l'on se fonde sur l'âge ou sur le temps écoulé depuis la fin des études.

5.3 La scolarisation augmente l'hypergamie et l'endogamie

Après cette analyse globale fondée sur l'écoulement du temps, nous nous tournons vers l'étude de l'effet de la scolarisation sur le choix du conjoint. L'hypothèse que nous cherchons à tester dans cette section stipule que la scolarisation augmente les chances de faire un choix hypergame ou endogame plutôt qu'hypogame : en effet, le système scolaire met Ego en contact avec des camarades dont le diplôme final sera plus proche du sien, mais aussi plus élevé que dans la population générale.

5.3.1 La scolarisation augmente l'hypergamie

Comme on pouvait s'y attendre, la scolarisation réduit chez les deux sexes les chances de se mettre en couple (Figure 6.9). Cet effet était légèrement plus fort chez les femmes que chez les hommes pour les cohortes les plus anciennes : ce point avait déjà été relevé par Maria Winkler-Dworak et Laurent Toulemon (2007, p. 293). Nous pouvons cependant noter que cette différence est surtout valable en ce qui concerne les chances de prendre un conjoint non diplômé, ou titulaire du certificat d'études primaires, du CAP, du BEP ou du BEPC, ce qui correspond pour ces générations à la grande majorité des cas. En revanche, contrairement à ces auteurs, nous observons une atténuation de l'effet chez les femmes au fil des cohortes, et une évolution en U inversé chez les hommes. Chez ces derniers, l'effet négatif de la scolarisation est réduit pour les générations nées autour de 1945, mais reprend de l'ampleur ensuite : de ce fait, parmi les cohortes les plus récentes, c'est désormais chez les hommes que la scolarisation présente l'effet restrictif le plus fort (Robert-Bobée et Mazuy, 2005, p. 14).

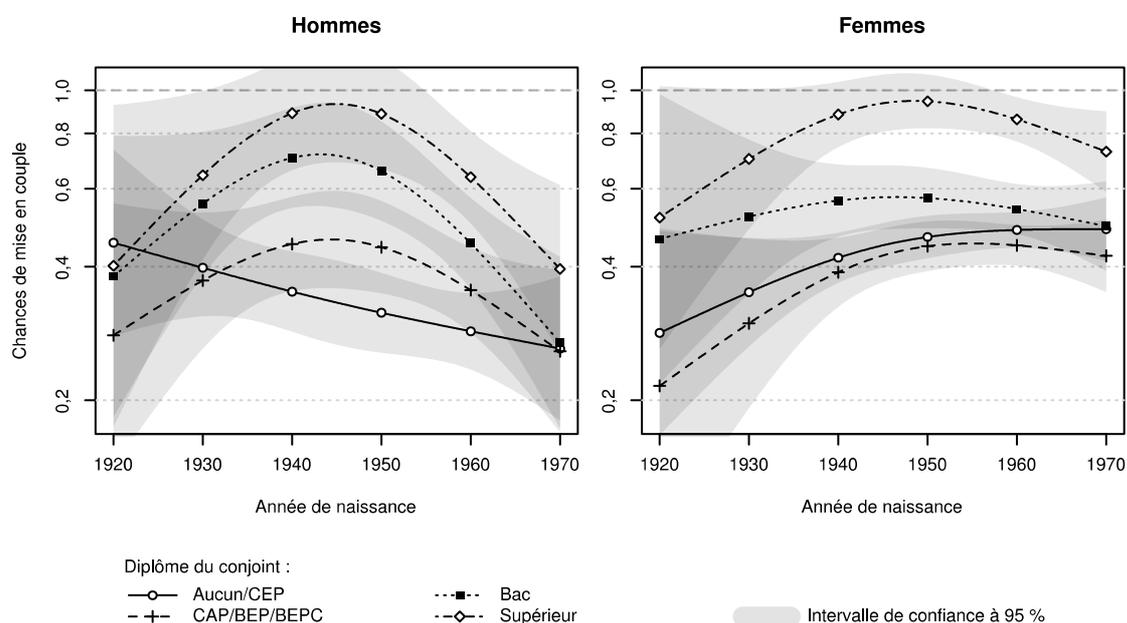
Au-delà de cet effet global sur les chances de se mettre en couple avec un conjoint de n'importe quel niveau de diplôme du conjoint, on observe chez les deux sexes que la scolarisation réduit fortement les chances de rencontrer un conjoint peu diplômé, mais moins celles de rencontrer un conjoint diplômé du supérieur. Chez les hommes et les femmes nés autour de 1945, la scolarisation n'a même pas d'effet significatif négatif sur les chances de mise en couple avec un diplômé du supérieur. Ainsi, puisque le modèle contrôle

l'effet du diplôme d'Ego, on peut conclure que se mettre en couple en cours de scolarité est dans l'ensemble un facteur important d'hypergamie, chez les hommes comme chez les femmes, validant ainsi l'hypothèse formulée plus haut. Ce phénomène a d'ailleurs été observé chez les deux sexes par W. Smeenk (1998, sect. 6.4.2.2) dans le cas des Pays-Bas.

Les chances de rencontrer un conjoint non diplômé, ou titulaire du seul certificat d'études, échappent en partie à ce schéma chez les femmes, en étant moins affectées par la scolarisation que les chances de rencontrer un titulaire du CAP, du BEP ou du BEPC. Cette observation est à rapprocher du fait qu'il est très rare pour les hommes de ce groupe de diplôme de se mettre en couple en cours de scolarité (Figure 6.2, p. 344) : le rôle des rencontres dans le cadre scolaire est donc nécessairement limité. Cette exception apparente semble ainsi confirmer le mécanisme envisagé.

Si l'effet négatif de la scolarisation semble donc globalement s'atténuer à mesure que s'élève le diplôme, il semble important de souligner que chez les hommes, l'écart est le plus important pour les cohortes nées entre 1940 et 1950, qui ont connu les conditions d'entrée sur le marché du travail les plus favorables, et qu'il s'est réduit pour les cohortes plus récentes. On retrouve ici le lien entre âge de première mise en couple et situation économique déjà abordé au chapitre précédent. On peut donc avancer que deux effets en partie contradictoires sont associés à la scolarisation. D'une part, chez les deux sexes, l'importance des rencontres dans le cadre scolaire implique que la scolarisation augmente le niveau de diplôme des conjoints de la minorité d'individus qui entre en cohabitation à cette période de la vie. Mais ce phénomène semble s'atténuer, en particulier chez les hommes, lorsque la situation économique est peu favorable : dans ce cas, la mise en couple en cours de scolarité est rendue très improbable, laissant moins de place à des différences selon le diplôme du conjoint, qui ne restent de ce fait visibles que pour les conjoints diplômés du supérieur.

Figure 6.9 : Effet de la scolarisation sur les chances de rencontrer un conjoint d'un niveau de diplôme donné par rapport à celles de rencontrer un non diplômé



Référence : 1 pour un conjoint de ce niveau de diplôme pendant la vie active

Le modèle 1 utilisé ici contrôle l'effet du diplôme final, de l'origine sociale et de l'âge au sein de chaque cohorte de naissance.

Lecture : Un homme né en 1950 voit ses chances de se mettre en couple avec une diplômée du supérieur diminuer de 10 % (multiplication par 0,9) lorsqu'il est scolarisé par rapport à celles qu'il aurait pendant la vie active ; ses chances de rencontrer une non diplômée diminuent de 70 % (multiplication par 0,3). La scolarisation multiplie donc par 3 (0,9/0,3) ses chances de rencontrer une diplômée du supérieur plutôt qu'une non diplômée.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France et n'ayant pas connu de séparation

5.3.2 La scolarisation augmente l'endogamie

Outre un effet sur le niveau absolu du diplôme du conjoint rencontré, les travaux existants se sont principalement interrogés sur l'effet de la scolarisation sur l'endogamie. Dans le cas de la France, M. Vanderschelden (2006b) avait déjà mesuré, à partir des mêmes données mais avec une méthode différente, une association positive entre les deux phénomènes. Nous observons bien cet effet (Figure 6.10), mais son ampleur est relativement limitée.

Chez les hommes, l'effet est faiblement significatif pour les cohortes d'avant-guerre⁴⁴, quoique le coefficient soit systématiquement positif et de l'ordre d'une augmentation des chances de 20 % à 40 %. La faible significativité est cohérente avec la très faible proportion de personnes ayant cohabité en cours d'études dans ces générations (Figure 6.2, p. 344). L'effet devient significatif et semble s'accroître légèrement, atteignant 40 %, pour les cohortes les plus récentes ; néanmoins, étant donné l'amplitude des intervalles de confiance, une stabilité au fil des cohortes est très plausible.

Chez les femmes, l'effet apparaît plus faible, de l'ordre de 10 % à 20 %, et ne varie quasiment pas au fil des cohortes ; comme chez les hommes, sa significativité est plus faible pour les cohortes les plus anciennes. Si les intervalles de confiance sont trop larges pour nous permettre d'affirmer avec certitude que l'effet est plus fort chez les hommes, ce résultat semble cohérent avec le fait que chez les femmes, les premières mises en couple en cours de scolarité sont bien plus fréquentes que les rencontres sur le lieu d'études mesurées par d'autres travaux : les femmes qui se mettent en couple en cours de scolarité rencontreraient plus souvent que les hommes leur conjoint en dehors du cadre scolaire, du fait notamment de l'écart d'âge en faveur de l'homme.

Dans l'ensemble, l'effet de la scolarisation sur l'endogamie apparaît relativement faible. L'augmentation de la fréquence des mises en couple en cours de scolarité que nous avons mise en évidence plus haut implique cependant que les effectifs concernés par ce phénomène ont très nettement crû au fil des cohortes, ce qui aurait dû logiquement entraîner une hausse de l'endogamie.

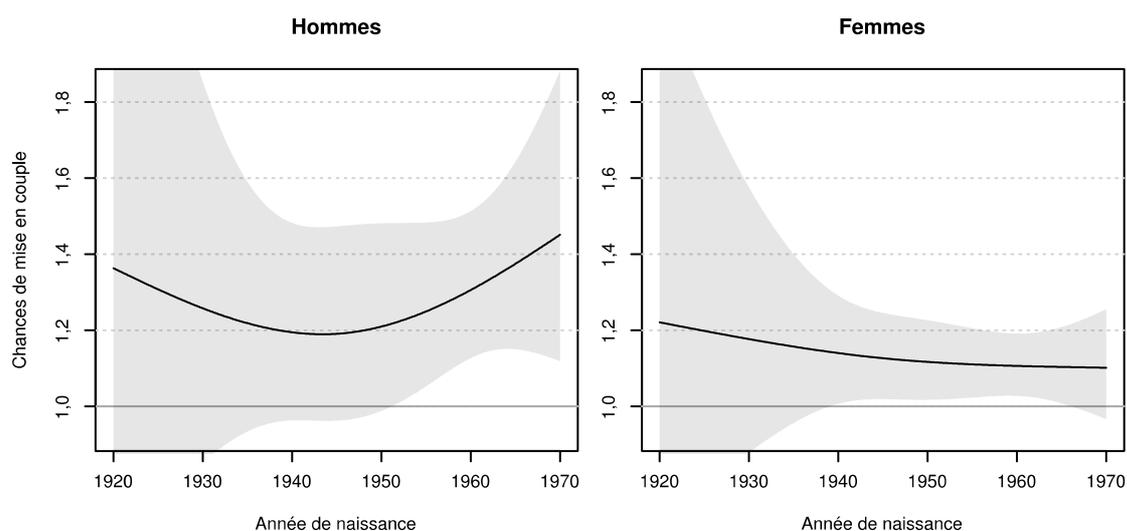
Contrairement à la scolarisation, le fait d'avoir terminé ses études, mais de ne jamais avoir occupé un emploi plus de trois mois, n'a pas d'effet significatif sur le choix du conjoint (Tableau 6.4, p. 387). Cela n'est pas étonnant dans la mesure où le nombre de mises en couple observées à cette période, relativement courte, est assez faible dans notre échantillon. Les intervalles de confiance ne permettent pas de conclure dans un sens plutôt que dans un autre.

Au total, les hypothèses proposées sont assez largement validées. La scolarisation réduit plus fortement les chances de former un couple avec un conjoint peu diplômé que fortement

44 La significativité apparaît plus élevée, chez les hommes comme chez les femmes, dans les modèles estimés séparément sur ces seules cohortes (*cf.* tableau 6.4).

diplômé, favorisant donc les chances relatives d'hypergamie. Elle renforce aussi les chances de former un couple endogame, quoique moins fortement qu'on aurait pu l'attendre. Nous allons maintenant tenter d'affiner ces résultats à l'aide de notre second modèle.

Figure 6.10 : Effet de la scolarisation sur les chances de rencontrer un conjoint de même diplôme que soi



Intervalle de confiance à 95 %

Référence : 1 pour rencontrer un conjoint de même diplôme que soi pendant la vie active

Le modèle 1 utilisé ici contrôle l'effet du diplôme final, de l'origine sociale et de l'âge au sein de chaque cohorte de naissance.

Lecture : Un homme né en 1970 a 40 % de chances en plus de rencontrer une conjointe possédant le même diplôme que lui pendant ses études que pendant la vie active.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France et n'ayant pas connu de séparation

5.4 La durée des études : une confirmation partielle de l'effet de la scolarisation

Plus que le fait d'être scolarisé au moment de la mise en couple, c'est le niveau d'études atteint une année donnée qui est censé, selon le mécanisme développé plus haut, renforcer l'endogamie et l'hypergamie. En effet, la population des élèves ou des étudiants est de plus en plus sélectionnée à mesure que l'on progresse dans le système scolaire. Le temps passé

dans le système scolaire est ici utilisé comme une indication de la position d'Ego dans ce processus de sélection.

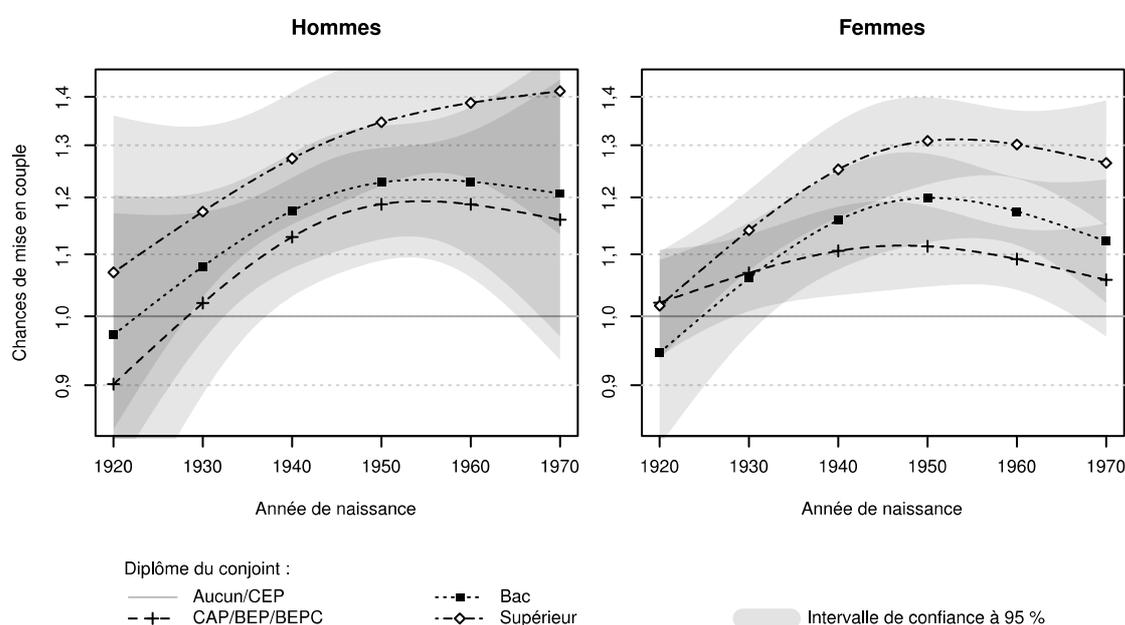
5.4.1 Le temps de scolarisation augmente l'hypergamie

Ainsi, si ce mécanisme fonctionne, on devrait s'attendre à observer une augmentation des chances de rencontrer un conjoint très diplômé avec le temps passé dans le système scolaire. De fait, le modèle 2 confirme et précise les résultats du modèle 1 concernant l'effet de la scolarisation. Rappelons que nos deux modèles contrôlent le diplôme finalement obtenu : l'effet observé ici est bien un effet de contexte, et non un simple effet de sélection⁴⁵. Plus encore, le fait qu'Ego soit scolarisé une année donnée est lui aussi contrôlé, si bien que le test de l'effet du temps de scolarisation est extrêmement exigeant du point de vue de la puissance statistique requise : on ne peut attendre une grande précision des coefficients et des intervalles de confiance.

Un effet significatif ressort cependant de ce modèle (Figure 6.11). Chez les hommes comme chez les femmes, une année d'études supplémentaire augmente les chances de rencontrer un conjoint très diplômé par rapport à celles de rencontrer un conjoint non diplômé ou titulaire du seul certificat d'études primaires. Cet effet est fort, puisqu'une année d'études supplémentaire ajoute jusqu'à 30 % de chances de rencontrer un diplômé du supérieur plutôt qu'un non diplômé ou un titulaire du seul certificat d'études. De même, l'augmentation des chances est d'autant plus forte que le diplôme du conjoint considéré est élevé. Notons néanmoins que cet effet ne s'observe pas, ou du moins pas aussi clairement, pour les générations nées dans les années 1920, pour lesquelles la fiabilité des données est cependant soumise à caution du fait de la mortalité différentielle aux âges avancés et de la faible proportion de diplômés en leur sein.

45 Par ailleurs, cet effet persiste dans un modèle contrôlant le diplôme de manière plus fine (en 7 catégories) que la classification retenue ici.

Figure 6.11 : Effet d'une année de scolarisation de plus sur les chances de rencontrer un conjoint d'un niveau de diplôme donné



Référence : 1 pour rencontrer un conjoint non diplômé en ayant terminé ses études à 14 ans

Le modèle 2 utilisé ici contrôle l'effet du diplôme final, de l'origine sociale, du statut et du temps écoulé depuis la fin des études au sein de chaque cohorte de naissance.

Lecture : Pour une femme née en 1950, à diplôme final contrôlé, chaque année d'études supplémentaire après 14 ans augmente de 30 % les chances de rencontrer un conjoint diplômé du supérieur, par rapport à celles de rencontrer un non diplômé ou un titulaire du seul certificat d'études ; soit 14 fois plus de chances ($1,3^{10}$) si elle prolonge ses études jusqu'à 24 ans.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France et n'ayant pas connu de séparation

5.4.2 Le temps de scolarisation n'augmente pas l'endogamie

Si l'effet positif de la scolarisation sur les chances de rencontrer un conjoint diplômé se trouve ainsi confirmé, qu'en est-il de celui sur les chances de former un couple endogame ? Notre second modèle ne donne pas de réponse claire à cette question. Aussi bien l'effet de la scolarisation que l'effet du temps passé dans le système scolaire apparaissent non significatifs (tableau 6.5, p. 391). Il semble ainsi difficile de distinguer empiriquement ces deux facteurs explicatifs, les deux variables étant fortement corrélées : les individus qui ont passé le plus de temps dans le système éducatif sont ceux qui se mettent le plus souvent en couple pendant leur scolarité. Comme nous l'avons expliqué dans la section précédente, le

caractère extrêmement exigeant du test réalisé par ce second modèle et l'ampleur des intervalles de confiance nous interdisent de conclure à une absence d'effet.

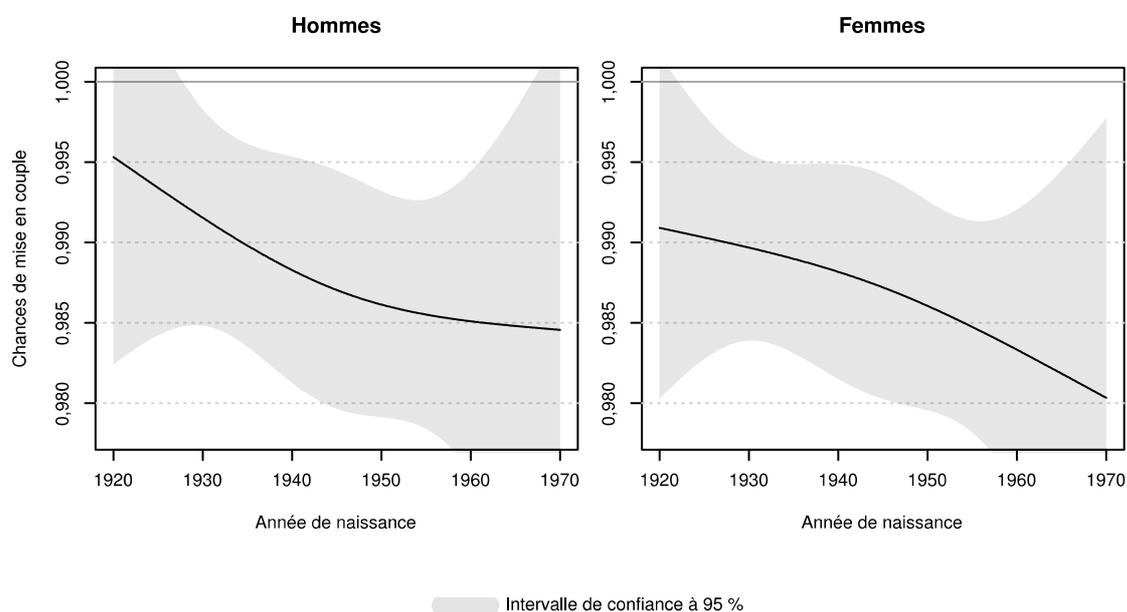
Notons que dans la littérature existante, W. Smeenk (1998, sect. 7.4.2.2) a observé un effet positif du nombre d'années passées dans le système scolaire sur les chances de rencontrer un conjoint de même niveau de diplôme que soi plutôt qu'un conjoint moins diplômé – quoique seulement chez les hommes. Mais son modèle ne contrôlait pas simultanément le fait d'être scolarisé une année donnée : aussi, il semble plausible que sa mesure capture, de manière indirecte, l'effet de la scolarisation.

5.5 Le temps écoulé depuis la fin des études diminue l'endogamie

Une dernière manière de préciser l'effet de la scolarisation sur l'endogamie consiste à mesurer l'effet du temps écoulé depuis la fin des études, censé signaler une plus grande hétérogénéité des fréquentations du point de vue du niveau d'études. Nous avons déjà observé plus haut qu'avec le temps, les chances de rencontrer un conjoint très diplômé augmentaient, plus fortement cependant chez les hommes que chez les femmes. Qu'en est-il des chances de rencontrer un conjoint de même diplôme que soi ?

La théorie du système scolaire comme marché conjugal se trouve ici renforcée, confirmant l'étude pionnière de R. Mare (1991) sur les États-Unis, ainsi que celle plus récente de M. Vanderschelden (2006b) sur la France. On observe chez les deux sexes un effet négatif significatif du temps écoulé depuis la fin des études sur les chances de former un couple endogame. Son ampleur est encore une fois limitée : environ 16 % de chances en moins 10 ans après la fin des études, le double 20 ans après. Néanmoins, son existence, dans un modèle qui contrôle par ailleurs le diplôme, l'origine sociale, le statut de scolarisation et le temps de scolarisation, constitue une confirmation forte de l'importance des réseaux liés aux études dans la formation de l'endogamie. Rappelons en effet que l'âge – retenu dans le premier modèle à la place du temps écoulé depuis la fin des études – ne présente aucun effet significatif sur l'endogamie : le mécanisme à l'œuvre semble bien être lié à la scolarité.

Figure 6.12 : Effet d'une année de plus depuis la fin des études sur les chances de former un couple endogame



Référence : 1 pour rencontrer un conjoint de même diplôme que soi l'année de fin d'études

Le modèle 2 utilisé ici contrôle l'effet du diplôme final, de l'origine sociale, du statut et du temps de scolarisation au sein de chaque cohorte de naissance.

Lecture : Pour une femme née en 1950, une année supplémentaire après la fin des études diminue de 1,5 % les chances de rencontrer un conjoint de même niveau de diplôme ; soit 14 % ($1 - 0,985^{10}$) de chances de moins 10 ans après la fin des études.

Source : enquête Étude de l'histoire familiale 1999 (Insee-Ined)

Champ : premiers couples des individus nés en France et n'ayant pas connu de séparation

6 Conclusion

Au terme de cette analyse, les hypothèses avancées par les théories du système éducatif comme marché conjugal, ainsi que celles que nous avons élaborées concernant les effets de la structure de la population des célibataires à un âge donné apparaissent largement validées. Nos résultats confirment les (rares) travaux existants dont la méthodologie nous semble valide. Résumons-les succinctement.

Tout d'abord, à diplôme final, origine sociale et scolarisation contrôlés, les chances de rencontrer un conjoint fortement diplômé (et donc de réaliser un choix hypergame) augmentent fortement avec l'âge (mais aussi avec le temps écoulé depuis la fin des études) chez les deux sexes, quoique moins nettement chez les femmes nées avant guerre ; on

observe une convergence très nette entre hommes et femmes au fil des générations. Ces évolutions suivent de manière assez proche celles de la structure par diplôme de la population des célibataires du sexe opposé. Ensuite, à diplôme final, origine sociale et âge contrôlés, la scolarisation augmente chez les deux sexes à la fois les chances d'endogamie et d'hypergamie, alors que le temps écoulé depuis la fin des études réduit les chances d'endogamie. Ces deux effets sur l'endogamie s'avèrent cependant d'ampleur assez limitée, ce qui semble pouvoir s'expliquer par le fait que les mises en couple en cours de scolarité ne renforcent véritablement l'endogamie que dans le cas, assez rare, où les deux conjoints sont scolarisés. Finalement, le temps passé dans le système scolaire ne présente pas d'effet sur l'endogamie au-delà de celui de la scolarisation, mais il augmente nettement les chances d'hypergamie.

Ces résultats apportent des réponses paradoxales aux interrogations formulées plus haut. Concernant l'affaiblissement de l'homogamie d'éducation que nous avons mis en évidence dans la première partie, nous constatons que celui-ci s'est déroulé en dépit du fait que la massification scolaire et l'augmentation de la fréquence des mises en couple en cours d'études entraînent mécaniquement une augmentation – limitée mais réelle – de l'endogamie. Ce paradoxe renforce la signification sociologique des évolutions temporelles établies en première partie, qui ne peuvent s'expliquer par des mécanismes directs tenant aux circonstances immédiates des rencontres amoureuses : on semble donc assister à un mouvement de fond de moindre détermination sociale du choix du conjoint.

Concernant la question de l'inversion de l'hypergamie féminine, nous observons bien un rapprochement entre hommes et femmes des mécanismes déterminant la mise en couple et le choix du conjoint, mais ce rapprochement reste inachevé, et contraste avec l'inversion de l'hypergamie absolue et relative que nous avons mise en évidence. En revanche, la fragilité de la norme d'hypergamie féminine se trouve encore une fois confirmée par la tendance des hommes âgés à choisir une conjointe plus diplômée, alors que leur position sur le marché conjugal leur permettrait de réaliser un choix hypogame. Si l'inversion de l'hypergamie de diplôme n'a pas trouvé ici son explication, les analyses de ce chapitre ont renforcé le constat d'un affaiblissement des différences genrées dans la mise en couple : les évolutions de l'âge de première mise en couple et du taux de célibat définitif par diplôme ont provoqué un rapprochement des profils de choix du conjoint par âge des hommes et des

femmes. Néanmoins, les facteurs économiques demeurent plus déterminants pour la mise en couple des hommes que pour celle des femmes, comme le montre l'effet négatif plus important de la scolarisation sur les chances de mise en couple chez les premiers par rapport aux secondes. Cette nuance confirme celle déjà formulée au chapitre précédent.

Pour finir, si l'importance de la structure du marché conjugal pour le choix du conjoint est apparue très clairement, on peut se demander si des facteurs de l'évolution de cette structure avec l'âge peuvent à leur tour être identifiés. Nous l'avons vu, celle-ci dépend de l'âge de mise en couple et du taux de célibat définitif par diplôme. Il serait donc intéressant de comprendre si ces deux dimensions dépendent uniquement de facteurs exogènes au marché conjugal (durée de la scolarisation, état du marché du travail, normes sociales), ou si le fonctionnement de ce dernier influe en retour sur le rythme de mise en couple et le célibat. Il est en effet possible que la composition même de la population des célibataires du sexe opposé influe sur le rythme de mise en couple des individus, par exemple en générant une attente plus longue sur le marché conjugal (Oppenheimer, 1988). De même, le célibat définitif peut être considéré non pas seulement comme le refus des intéressés ou de leurs conjoints potentiels de se mettre en couple, mais aussi comme un désajustement sur le marché conjugal – par exemple comme un décalage de calendriers entre groupes (*cf.* chapitre précédent). Ainsi, la présente étude semble pouvoir être prolongée par le développement de techniques qui, contrôlant l'effet de la composition de la population des célibataires du sexe opposé à chaque âge, permettraient de mesurer les écarts à ce modèle purement structurel, et leur évolution au cours du temps.

7 Appendice : Résultats détaillés des modèles à risques concurrents estimés séparément sur deux cohortes

Pour permettre au lecteur d'accéder aux résultats non repris sous forme graphique – et notamment les résultats non significatifs – nous présentons ici les coefficients détaillés des modèles à risques concurrents estimés séparément sur deux cohortes de naissance (1925-1944 et 1945-1964). Nous résumons rapidement l'interprétation des coefficients qui ne sont pas mentionnés dans le corps du texte.

Du fait de la présence systématique d'interactions entre les variables explicatives, d'une part, et entre le diplôme du conjoint et la nature endogame de l'union, d'autre part, les effets

principaux des variables décrivent uniquement les chances de former un couple avec un conjoint non diplômé – et non l'effet moyen dans l'ensemble de l'échantillon. Notons par ailleurs que les effets principaux de l'âge et du diplôme d'Ego, ainsi que l'interaction entre ces deux variables, ne s'expriment pas sous la forme de coefficients, mais donnent lieu à l'estimation de courbes de séjour séparées, non reprises ici.

Les coefficients correspondant au niveau de diplôme du conjoint, de plus en plus négatifs à mesure que le diplôme s'élève, signalent que les chances de mise en couple avec un conjoint non diplômé ou titulaire du seul certificat d'études primaires (catégorie de référence pour le diplôme du conjoint) diminuent à mesure que le diplôme d'Ego augmente.

Concernant la catégorie socioprofessionnelle du père, on observe comme attendu que les individus dont le père appartient à une catégorie socioprofessionnelle autre que celle des ouvriers ont moins de chances de se mettre en couple avec un conjoint peu diplômé que les individus issus de ce dernier groupe (on ne note que deux exceptions, dans lesquelles le coefficient est négatif mais non significatif). L'effet est le plus fort pour les enfants de cadres et professions intellectuelles supérieures.

Le fait d'être encore scolarisé ou de ne jamais avoir occupé d'emploi pendant plus de trois mois diminue les chances d'entrer en cohabitation avec un conjoint non diplômé ou titulaire du certificat d'études. Dans le modèle 2, l'effet principal de la modalité « scolarisé » correspond à un individu scolarisé à l'âge de 14 ans (années de scolarisation fixées à 0), et la significativité est nettement réduite du fait du contrôle du temps de scolarisation.

L'interaction entre diplôme d'Ego et diplôme du conjoint permet de contrôler l'association entre différents groupes de diplôme, qui n'évolue pas au cours de la vie des individus, de manière à se concentrer sur l'effet des variables biographiques. Elle indique par exemple que les diplômés du supérieur ont des chances de rencontrer un conjoint d'un groupe de diplôme donné (par rapport à celles des non diplômés) d'autant plus fortes que ce diplôme est élevé. Soulignons que les coefficients correspondent ici aux chances de rencontre relatives à celles des non diplômés ou titulaires d'un certificat d'études : un coefficient positif indique de plus grandes chances de rencontre, mais le coefficient le plus élevé ne correspond pas au diplôme du conjoint le plus fréquemment associé à ce groupe. Cette double interaction ne peut s'interpréter complètement sans la combiner avec l'effet

principal du diplôme du conjoint, ainsi qu'avec les courbes de séjour spécifiques à chaque groupe de diplôme d'Ego (strates).

L'interaction entre PCS du père et diplôme du conjoint indique elle aussi qu'à autre caractéristiques contrôlées, l'origine sociale a un effet sur le choix du conjoint : les individus issus de toute autre catégorie socioprofessionnelle que celle des ouvriers ont par exemple plus de chances de rencontrer un conjoint titulaire du baccalauréat ou d'un diplôme du supérieur (des exceptions peuvent être notées chez les agriculteurs et la catégorie « autres », avec des coefficients négatifs et significatifs) ; les résultats sont moins nets concernant le niveau CAP/BEP/BEPC.

Tableau 6.4 : Résultats détaillés du modèle 1 appliqué séparément à deux générations

	Hommes				Femmes			
	1925-44		1945-64		1925-44		1945-64	
Diplôme du conjoint (réf. : Aucun/CEP = 0)								
CAP/BEP/BEPC	-2,32 ***	(0,23)	-1,11 ***	(0,15)	-1,29 ***	(0,09)	-0,32 ***	(0,07)
Bac	-5,12 ***	(0,34)	-3,80 ***	(0,19)	-3,74 ***	(0,14)	-3,19 ***	(0,10)
Supérieur	-8,26 ***	(0,58)	-5,65 ***	(0,22)	-5,39 ***	(0,20)	-4,89 ***	(0,12)
PCS du père d'Ego (réf. : Ouvrier = 0)								
Agriculteur	-0,33 ***	(0,05)	-0,20 ***	(0,04)	-0,04	(0,06)	-0,10 **	(0,04)
Indépendant	-0,24 ***	(0,05)	-0,27 ***	(0,06)	-0,28 ***	(0,06)	-0,47 ***	(0,05)
Cadre	-0,60 ***	(0,10)	-0,64 ***	(0,11)	-0,83 ***	(0,11)	-0,86 ***	(0,10)
Intermédiaire	-0,18 *	(0,08)	-0,46 ***	(0,07)	-0,51 ***	(0,08)	-0,38 ***	(0,06)
Employé	-0,13 *	(0,05)	-0,23 ***	(0,05)	-0,19 ***	(0,05)	-0,14 ***	(0,04)
Autre	-0,50 ***	(0,08)	-0,27 ***	(0,07)	-0,63 ***	(0,09)	-0,53 ***	(0,07)
Statut d'Ego (réf. : A déjà occupé un emploi = 0)								
Jamais travaillé	-0,84 ***	(0,22)	-1,38 ***	(0,28)	-0,19 .	(0,10)	-0,31 **	(0,10)
Scolarisé	-1,14 ***	(0,14)	-1,11 ***	(0,11)	-1,01 ***	(0,10)	-0,75 ***	(0,05)

	Hommes				Femmes			
	1925-44		1945-64		1925-44		1945-64	
Diplôme d'Ego × Diplôme du conjoint (réf. : Aucun × Aucun = 0)								
CAP/BEP/BEPC × CAP/BEP/BEPC	0,98 **	(0,31)	0,82 ***	(0,18)	1,17 ***	(0,16)	0,89 ***	(0,10)
CAP/BEP/BEPC × Bac	1,24 ***	(0,19)	1,10 ***	(0,11)	1,67 ***	(0,11)	1,36 ***	(0,08)
CAP/BEP/BEPC × Supérieur	1,40 ***	(0,22)	1,29 ***	(0,12)	2,19 ***	(0,12)	1,73 ***	(0,09)
Bac × CAP/BEP/BEPC	1,66 ***	(0,18)	1,41 ***	(0,12)	1,08 ***	(0,12)	1,14 ***	(0,07)
Bac × Bac	2,89 ***	(0,34)	2,34 ***	(0,21)	2,85 ***	(0,20)	2,23 ***	(0,13)
Bac × Supérieur	3,16 ***	(0,24)	2,72 ***	(0,14)	3,41 ***	(0,15)	2,80 ***	(0,10)
Supérieur × CAP/BEP/BEPC	1,96 ***	(0,19)	1,68 ***	(0,14)	1,16 ***	(0,15)	1,35 ***	(0,09)
Supérieur × Bac	3,41 ***	(0,22)	2,71 ***	(0,15)	2,62 ***	(0,18)	2,50 ***	(0,11)
Supérieur × Sup.	4,73 ***	(0,37)	4,21 ***	(0,23)	4,54 ***	(0,23)	4,09 ***	(0,15)
PCS du père d'Ego × Diplôme du conjoint (réf. : Ouvrier × Aucun/CEP = 0)								
Agriculteur × CAP/BEP/BEPC	-0,04	(0,06)	0,07	(0,05)	-0,57 ***	(0,06)	0,04	(0,04)
Agriculteur × Bac	0,18	(0,11)	0,22 **	(0,07)	-0,53 ***	(0,10)	0,15 *	(0,06)
Agriculteur × Supérieur	0,21	(0,13)	0,21 **	(0,07)	-0,26 *	(0,11)	0,10 .	(0,06)
Indépendant × CAP/BEP/BEPC	0,26 ***	(0,07)	0,23 ***	(0,07)	0,10 .	(0,06)	0,31 ***	(0,06)
Indépendant × Bac	0,62 ***	(0,11)	0,43 ***	(0,08)	0,32 ***	(0,09)	0,66 ***	(0,07)
Indépendant × Supérieur	0,42 ***	(0,12)	0,54 ***	(0,08)	0,76 ***	(0,10)	0,84 ***	(0,07)
Cadre × CAP/BEP/BEPC	0,50 ***	(0,13)	0,18	(0,12)	0,23 .	(0,14)	0,29 **	(0,11)
Cadre × Bac	0,76 ***	(0,16)	0,77 ***	(0,13)	0,88 ***	(0,16)	0,84 ***	(0,12)
Cadre × Supérieur	0,98 ***	(0,17)	0,94 ***	(0,13)	1,47 ***	(0,15)	1,41 ***	(0,11)
Intermédiaire × CAP/BEP/BEPC	0,33 **	(0,10)	0,37 ***	(0,08)	0,39 ***	(0,09)	0,18 **	(0,07)
Intermédiaire × Bac	0,41 **	(0,14)	0,75 ***	(0,09)	0,71 ***	(0,12)	0,56 ***	(0,08)

	Hommes				Femmes			
	1925-44		1945-64		1925-44		1945-64	
Intermédiaire × Supérieur	0,53 ***	(0,15)	0,86 ***	(0,09)	1,07 ***	(0,12)	0,72 ***	(0,08)
Employé × CAP/BEP/BEPC	0,30 ***	(0,06)	0,14 *	(0,06)	0,14 *	(0,06)	0,01	(0,05)
Employé × Bac	0,36 ***	(0,11)	0,33 ***	(0,07)	0,36 ***	(0,09)	0,27 ***	(0,06)
Employé × Supérieur	0,34 **	(0,12)	0,38 ***	(0,07)	0,48 ***	(0,10)	0,35 ***	(0,06)
Autre × CAP/BEP/BEPC	0,03	(0,09)	-0,35 ***	(0,08)	-0,06	(0,09)	-0,09	(0,07)
Autre × Bac	-0,07	(0,19)	-0,20 .	(0,12)	-0,15	(0,16)	-0,02	(0,11)
Autre × Supérieur	0,14	(0,20)	-0,53 ***	(0,12)	0,58 ***	(0,14)	0,10	(0,10)
Statut d'Ego × Diplôme du conjoint (réf. : A déjà occupé un emploi × Aucun/CEP = 0)								
Jamais travaillé × CAP/BEP/BEPC	0,38	(0,27)	0,34	(0,31)	-0,31 **	(0,11)	-0,08	(0,09)
Jamais trav. × Bac	0,72 *	(0,29)	0,45	(0,34)	0,02	(0,17)	0,01	(0,14)
Jamais trav. × Sup.	0,87 **	(0,30)	1,13 ***	(0,31)	-0,04	(0,18)	-0,15	(0,16)
Scolarisé × CAP/BEP/BEPC	0,21	(0,17)	0,21 .	(0,12)	0,01	(0,13)	-0,07	(0,06)
Scolarisé × Bac	0,51 **	(0,18)	0,53 ***	(0,13)	0,29 .	(0,15)	0,15 *	(0,07)
Scolarisé × Sup.	0,78 ***	(0,21)	0,83 ***	(0,13)	0,76 ***	(0,14)	0,61 ***	(0,06)
Âge d'Ego × Diplôme du conjoint (réf. : Aucun/CEP = 0)								
Âge × CAP/BEP/BEPC	0,12 ***	(0,02)	0,14 ***	(0,02)	0,08 ***	(0,01)	0,08 ***	(0,01)
Âge × Bac	0,22 ***	(0,04)	0,28 ***	(0,02)	0,16 ***	(0,02)	0,20 ***	(0,02)
Âge × Supérieur	0,48 ***	(0,06)	0,45 ***	(0,03)	0,26 ***	(0,03)	0,36 ***	(0,02)
Âge' × CAP/BEP/BEPC	-0,09 ***	(0,02)	-0,12 ***	(0,02)	-0,08 ***	(0,01)	-0,09 ***	(0,01)
Âge' × Bac	-0,15 ***	(0,03)	-0,23 ***	(0,02)	-0,14 ***	(0,02)	-0,18 ***	(0,02)
Âge' × Supérieur	-0,31 ***	(0,05)	-0,33 ***	(0,02)	-0,22 ***	(0,02)	-0,29 ***	(0,02)

	Hommes				Femmes			
	1925-44		1945-64		1925-44		1945-64	
PCS du père d'Ego × Couple endogame (réf. : Ouvrier × Couple exogame = 0)								
Agriculteur × Endo.	0,06	(0,06)	0,01	(0,04)	0,23 ***	(0,06)	0,10 **	(0,03)
Indépendant × Endo.	0,02	(0,06)	0,01	(0,05)	-0,02	(0,06)	0,02	(0,04)
Cadre × Endo.	-0,01	(0,12)	0,09	(0,08)	-0,05	(0,10)	-0,06	(0,07)
Interméd. × Endo.	-0,03	(0,09)	-0,04	(0,06)	0,00	(0,08)	0,01	(0,05)
Employé × Endo.	-0,01	(0,06)	0,00	(0,05)	0,01	(0,05)	-0,02	(0,04)
Autre × Endo.	0,16 .	(0,09)	0,24 ***	(0,07)	0,11	(0,08)	0,16 **	(0,06)
Statut d'Ego × Couple endogame (réf. : A déjà occupé un emploi × Couple exogame = 0)								
Jamais trav. × Endo.	-0,31	(0,24)	0,03	(0,19)	-0,07	(0,10)	0,09	(0,08)
Scolarisé × Endo.	0,35 *	(0,17)	0,19 *	(0,09)	0,18 .	(0,10)	0,11 **	(0,04)
Âge d'Ego × Couple endogame (réf. : Couple exogame = 0)								
Âge × Endo.	0,03	(0,02)	0,01	(0,01)	0,00	(0,01)	0,01	(0,01)
Âge' × Endo.	-0,03	(0,02)	-0,01	(0,01)	-0,01	(0,01)	-0,01	(0,01)
Périodes	1 100 564		1 495 560		1 212 312		1 627 432	
Événements	19 368		29 956		23 912		42 577	
Degrés de liberté	60		60		60		60	
Concordance	80,58 %		70,95 %		78,74 %		73,44 %	
R ² de Nagelkerke	6,60 %		3,40 %		5,71 %		4,24 %	
Log-vraisemblance	-152 415		-257 821		-195 526		-361 186	
Log-vraisemblance du modèle vide	-161 739		-265 451		-205 561		-373 993	
AIC	304 951		515 762		391 171		722 492	
BIC	305 423		516 260		391 656		723 012	

Les coefficients présentés sont exprimés sur une échelle logarithmique et additive, qui présentent l'avantage d'être symétriques autour de zéro. Le passage à des coefficients multiplicateurs et à des pourcentages d'augmentation ou de diminution s'opère en prenant l'exponentielle de ces valeurs.

Âge et Âge' désignent les deux coefficients associés à la base des splines cubiques restreintes construite à partir de l'âge d'Ego (cf. Encadré 5.2, p. 308).

Tableau 6.5 : Résultats détaillés du modèle 2 appliqué séparément à deux générations

	Hommes				Femmes			
	1925-44		1945-64		1925-44		1945-64	
Diplôme du conjoint (réf. : Aucun/CEP = 0)								
CAP/BEP/BEPC	-1,81 ***	(0,07)	-0,38 ***	(0,06)	-0,91 ***	(0,05)	-0,02	(0,04)
Bac	-3,88 ***	(0,12)	-2,12 ***	(0,08)	-2,90 ***	(0,08)	-2,35 ***	(0,06)
Supérieur	-5,19 ***	(0,17)	-2,88 ***	(0,10)	-3,85 ***	(0,11)	-3,16 ***	(0,08)
PCS du père d'Ego (réf. : Ouvrier = 0)								
Agriculteur	-0,34 ***	(0,05)	-0,20 ***	(0,04)	-0,05	(0,06)	-0,10 **	(0,04)
Indépendant	-0,24 ***	(0,06)	-0,27 ***	(0,06)	-0,28 ***	(0,06)	-0,48 ***	(0,05)
Cadre	-0,61 ***	(0,10)	-0,65 ***	(0,11)	-0,81 ***	(0,11)	-0,88 ***	(0,10)
Intermédiaire	-0,17 *	(0,08)	-0,46 ***	(0,07)	-0,51 ***	(0,08)	-0,40 ***	(0,06)
Employé	-0,13 *	(0,05)	-0,23 ***	(0,05)	-0,19 ***	(0,05)	-0,15 ***	(0,04)
Autre	-0,52 ***	(0,08)	-0,29 ***	(0,07)	-0,64 ***	(0,09)	-0,55 ***	(0,07)
Statut d'Ego (réf. : A déjà occupé un emploi = 0)								
Jamais travaillé	-0,73 ***	(0,22)	-1,24 ***	(0,28)	-0,15	(0,11)	-0,24 *	(0,10)
Scolarisé	-0,35	(0,50)	-0,05	(0,25)	-0,15	(0,23)	-0,22 *	(0,11)
Années de scolarisation d'Ego								
Années	-0,07	(0,07)	-0,13 **	(0,05)	-0,15 **	(0,05)	-0,08 **	(0,03)
Années depuis la fin des études d'Ego								
Années	0,02 **	(0,01)	0,02 **	(0,01)	0,03 **	(0,01)	0,01	(0,01)
Diplôme d'Ego × Diplôme du conjoint (réf. : Aucun × Aucun = 0)								
CAP/BEP/BEPC × CAP/BEP/BEPC	1,60 ***	(0,10)	1,07 ***	(0,08)	1,37 ***	(0,08)	1,14 ***	(0,06)
CAP/BEP/BEPC × Bac	1,54 ***	(0,11)	1,24 ***	(0,08)	1,80 ***	(0,08)	1,57 ***	(0,06)
CAP/BEP/BEPC × Supérieur	1,67 ***	(0,16)	1,44 ***	(0,09)	2,32 ***	(0,10)	1,99 ***	(0,08)
Bac × CAP/BEP/BEPC	2,01 ***	(0,10)	1,58 ***	(0,09)	1,21 ***	(0,10)	1,32 ***	(0,06)
Bac × Bac	3,52 ***	(0,15)	2,66 ***	(0,12)	3,12 ***	(0,13)	2,67 ***	(0,09)
Bac × Supérieur	3,57 ***	(0,17)	3,03 ***	(0,11)	3,64 ***	(0,13)	3,28 ***	(0,09)
Supérieur × CAP/BEP/BEPC	2,37 ***	(0,12)	1,87 ***	(0,11)	1,27 ***	(0,14)	1,53 ***	(0,08)

	Hommes				Femmes			
	1925-44		1945-64		1925-44		1945-64	
Supérieur × Bac	3,88 ***	(0,15)	3,02 ***	(0,12)	2,84 ***	(0,15)	2,88 ***	(0,10)
Supérieur × Sup.	5,69 ***	(0,19)	4,85 ***	(0,14)	4,92 ***	(0,16)	4,87 ***	(0,11)
PCS du père d'Ego × Diplôme du conjoint (réf. : Ouvrier × Aucun/CEP = 0)								
Agriculteur × CAP/BEP/BEPC	-0,03	(0,06)	0,08	(0,05)	-0,56 ***	(0,06)	0,04	(0,04)
Agriculteur × Bac	0,20 .	(0,11)	0,23 ***	(0,07)	-0,52 ***	(0,10)	0,15 *	(0,06)
Agriculteur × Supérieur	0,23 .	(0,13)	0,22 **	(0,07)	-0,23 *	(0,11)	0,11 .	(0,06)
Indépendant × CAP/BEP/BEPC	0,27 ***	(0,07)	0,23 ***	(0,07)	0,12 .	(0,06)	0,32 ***	(0,06)
Indépendant × Bac	0,65 ***	(0,11)	0,45 ***	(0,08)	0,35 ***	(0,09)	0,69 ***	(0,07)
Indépendant × Supérieur	0,52 ***	(0,12)	0,58 ***	(0,08)	0,80 ***	(0,10)	0,89 ***	(0,07)
Cadre × CAP/BEP/BEPC	0,53 ***	(0,13)	0,20	(0,12)	0,24 .	(0,14)	0,30 **	(0,11)
Cadre × Bac	0,81 ***	(0,16)	0,81 ***	(0,13)	0,91 ***	(0,16)	0,89 ***	(0,12)
Cadre × Supérieur	1,10 ***	(0,18)	1,03 ***	(0,13)	1,51 ***	(0,15)	1,50 ***	(0,11)
Intermédiaire × CAP/BEP/BEPC	0,33 ***	(0,10)	0,38 ***	(0,08)	0,40 ***	(0,09)	0,19 **	(0,07)
Intermédiaire × Bac	0,43 **	(0,14)	0,78 ***	(0,09)	0,73 ***	(0,12)	0,58 ***	(0,08)
Intermédiaire × Supérieur	0,62 ***	(0,15)	0,91 ***	(0,09)	1,11 ***	(0,12)	0,78 ***	(0,08)
Employé × CAP/BEP/BEPC	0,31 ***	(0,06)	0,15 **	(0,06)	0,14 *	(0,06)	0,01	(0,05)
Employé × Bac	0,38 ***	(0,11)	0,35 ***	(0,07)	0,37 ***	(0,09)	0,29 ***	(0,06)
Employé × Supérieur	0,39 **	(0,12)	0,42 ***	(0,07)	0,51 ***	(0,10)	0,38 ***	(0,06)
Autre × CAP/BEP/BEPC	0,04	(0,09)	-0,34 ***	(0,08)	-0,05	(0,09)	-0,08	(0,07)
Autre × Bac	-0,02	(0,19)	-0,17	(0,12)	-0,12	(0,16)	0,01	(0,11)
Autre × Supérieur	0,19	(0,20)	-0,48 ***	(0,12)	0,65 ***	(0,14)	0,16	(0,10)

	Hommes				Femmes			
	1925-44		1945-64		1925-44		1945-64	
Statut d'Ego × Diplôme du conjoint (réf. : A déjà occupé un emploi × Aucun/CEP = 0)								
Jamais travaillé × CAP/BEP/BEPC	0,35	(0,27)	0,23	(0,31)	-0,35 **	(0,11)	-0,14	(0,09)
Jamais trav. × Bac	0,65 *	(0,29)	0,26	(0,34)	-0,06	(0,17)	-0,12	(0,14)
Jamais trav. × Sup.	0,80 **	(0,30)	0,87 **	(0,30)	-0,20	(0,18)	-0,37 *	(0,16)
Scolarisé × CAP/BEP/BEPC	-0,07	(0,56)	-0,60 *	(0,28)	-0,66 **	(0,25)	-0,45 ***	(0,11)
Scolarisé × Bac	-0,22	(0,56)	-0,61 *	(0,30)	-0,33	(0,34)	-0,83 ***	(0,14)
Scolarisé × Supérieur	-1,23 *	(0,61)	-1,18 ***	(0,30)	-1,03 ***	(0,28)	-0,98 ***	(0,13)
Années de scolarisation × Diplôme du conjoint (réf. : Aucun/CEP = 0)								
Années × CAP/BEP/BEPC	0,03	(0,08)	0,12 *	(0,05)	0,15 *	(0,06)	0,07 *	(0,03)
Années × Bac	0,08	(0,08)	0,15 **	(0,05)	0,12 .	(0,07)	0,16 ***	(0,03)
Années × Supérieur	0,23 **	(0,08)	0,25 ***	(0,05)	0,29 ***	(0,06)	0,21 ***	(0,03)
Années depuis la fin des études × Diplôme du conjoint (réf. : Aucun/CEP = 0)								
Années × CAP/BEP/BEPC	0,01 *	(0,00)	0,01	(0,00)	0,00	(0,00)	0,00	(0,00)
Années × Bac	0,02 *	(0,01)	0,01 .	(0,01)	0,00	(0,01)	0,02 **	(0,01)
Années × Supérieur	0,06 ***	(0,01)	0,04 ***	(0,01)	0,00	(0,01)	0,04 ***	(0,01)
PCS du père d'Ego × Couple endogame (réf. : Ouvrier × Couple exogame = 0)								
Agriculteur × Endo.	0,07	(0,06)	0,02	(0,04)	0,24 ***	(0,06)	0,10 **	(0,03)
Indépendant × Endo.	0,03	(0,06)	0,02	(0,05)	-0,02	(0,06)	0,02	(0,04)
Cadre × Endo.	-0,01	(0,12)	0,09	(0,08)	-0,06	(0,10)	-0,07	(0,07)
Interméd. × Endo.	-0,02	(0,09)	-0,03	(0,06)	0,00	(0,08)	0,01	(0,05)
Employé × Endo.	-0,01	(0,06)	0,00	(0,05)	0,01	(0,05)	-0,01	(0,04)
Autre × Endo.	0,18 *	(0,09)	0,25 ***	(0,07)	0,11	(0,08)	0,16 **	(0,06)

	Hommes		Femmes	
	1925-44	1945-64	1925-44	1945-64
Statut d'Ego × Couple endogame (réf. : A déjà occupé un emploi × Couple exogame = 0)				
Jamais trav. × Endo.	-0,37 (0,23)	0,05 (0,19)	-0,09 (0,11)	0,07 (0,08)
Scolarisé × Endo.	-0,03 (0,44)	-0,15 (0,18)	0,06 (0,21)	-0,11 (0,08)
Années de scolarisation × Couple endogame (réf. : Couple exogame = 0)				
Années × Endo.	0,02 (0,06)	0,03 (0,03)	-0,02 (0,04)	0,02 (0,02)
Années depuis la fin des études × Couple endogame (réf. : Couple exogame = 0)				
Années × Endo.	-0,02 *** (0,00)	-0,01 * (0,00)	-0,01 ** (0,00)	-0,01 *** (0,00)
Périodes	1 100 564	1 495 560	1 212 312	1 627 432
Événements	19 368	29 956	23 912	42 577
Degrés de liberté	62	62	62	62
Concordance	80,52 %	70,88 %	78,74 %	73,21 %
R ² de Nagelkerke	6,53 %	3,29 %	5,67 %	4,15 %
Log-vraisemblance	-152 509	-258 073	-195 595	-361 444
Log-vraisemblance du modèle vide	-161 739	-265 451	-205 561	-373 993
AIC	305 142	516 270	391 313	723 012
BIC	305 630	516 785	391 815	723 549

Les coefficients présentés sont exprimés sur une échelle logarithmique et additive, qui présentent l'avantage d'être symétriques autour de zéro. Le passage à des coefficients multiplicateurs et à des pourcentages d'augmentation ou de diminution s'opère en prenant l'exponentielle de ces valeurs.

PARTIE III

VARIATIONS ET DÉTERMINANTS DE L'HOMOGAMIE ET DE L'HYPERGAMIE EN EUROPE

CHAPITRE 7

LA STRUCTURE ET LE NIVEAU DE L'HOMOGAMIE ET DE L'HYPERGAMIE SOCIOÉCONOMIQUE ET D'ÉDUCATION : DES DIFFÉRENCES MARQUÉES ENTRE ET À L'INTÉRIEUR DES PAYS

Les analyses conduites dans les deux premières parties à propos du cas français ont mis en évidence l'affaiblissement aussi bien de l'homogamie que de l'hypergamie – bien qu'avec des nuances –, en cohérence avec les théories de la modernisation. Toutefois, une étude précise des déterminants du choix du conjoint ne saurait se fonder sur un seul pays : une plus grande diversité de configurations est nécessaire pour tenter de démêler les influences des différents facteurs. C'est pourquoi cette troisième partie élargit le champ d'étude aux pays de l'Union européenne – et plus précisément à 64 régions qui les composent –, en passant d'une comparaison dans le temps à une comparaison dans l'espace.

Le champ géographique retenu présente des avantages comme des limites. Ces dernières tiennent à la relative homogénéité, à l'échelle mondiale ou historique, des sociétés européennes contemporaines : la variété de leurs structures sociales et de leurs niveaux de développement est relativement restreinte, et ne permet pas de tenter de mettre à l'épreuve dans leur ensemble des théories aussi générales que celles de la modernisation. Les analyses qui suivent ne peuvent ainsi porter que sur les déterminants de l'homogamie et de l'hypergamie à un stade avancé de la modernisation. Comme nous le verrons plus bas, ces remarques ne doivent cependant pas nous conduire à sous-estimer l'ampleur des différences entre pays, et plus encore entre régions, de l'Union européenne, qui sont largement suffisantes pour étudier les déterminants du choix du conjoint. Les structures sociales demeurent fortement contrastées entre Est et Ouest aussi bien qu'entre Nord et Sud (Breen, 2004 ; Róbert, 2010). Les modèles familiaux, et notamment les modalités du départ du domicile parental et de première mise en couple, varient eux aussi fortement, en partie selon

les mêmes oppositions Est-Ouest et Nord-Sud (Reher, 1998 ; Billari, Philipov et Baizán, 2001 ; Sobotka et Toulemon, 2008 ; Puur *et al.*, 2012).

À l'inverse, les avantages d'une analyse restreinte à l'Europe sont multiples. Sur le plan théorique, l'homogénéité est aussi un atout (Smits et Park, 2009, p. 229), dans la mesure où elle autorise des mises en équivalence dont la validité serait délicate à déterminer dans le cadre d'une comparaison de pays aux niveaux de développement très différents. Sur le plan pratique, des données riches et de qualité sont disponibles à l'échelle européenne, du fait entre autres de l'harmonisation institutionnelle mise en œuvre sous l'égide de l'institut statistique européen Eurostat, mais aussi d'enquêtes conduites par des réseaux indépendants, comme l'Enquête sociale européenne (*European Social Survey*). Cette richesse des données nous permet de décrire les différences régionales internes à chaque pays, de manière à la fois à évaluer leur homogénéité et à renforcer les variations entre unités d'analyse. Ainsi, quand plusieurs études existantes ont retenu une approche extensive fondée sur un large échantillon de pays très divers, nous retiendrons ici une démarche intensive fondée sur un champ plus restreint mais des informations nettement plus détaillées. Nous serons ainsi en mesure de décrire simultanément les dimensions éducative et socioéconomique du choix du conjoint, ce qui n'a jamais été réalisé dans la littérature, et d'évaluer (au chapitre suivant) l'effet de déterminants relativement fins restés jusque-là peu explorés.

Nous commençons par présenter les études comparatives existantes et les débats qui persistent concernant les déterminants de l'homogamie. Nous introduisons ensuite les données de l'Enquête européenne sur les forces de travail (*European Union Labour Force Survey*, EU-LFS) mobilisées ici, ainsi que les difficultés liées à l'harmonisation entre pays de la définition du niveau d'éducation et du groupe socioéconomique. Après avoir développé la stratégie de modélisation retenue, nous présentons les variations entre régions européennes de l'homogamie et de l'hypergamie, dans leurs mesures absolue puis relative. Des différences très marquées apparaissent du point de vue du niveau comme de la structure de l'homogamie, ainsi que du sens (féminine ou masculine) et de l'intensité de l'hypergamie. Si des régularités peuvent être relevées, les composantes étudiées ne sont que partiellement corrélées entre elles. Au total, ce premier aperçu indique que l'homogamie décroît avec le

niveau de développement ; cette relation ne se retrouve cependant pas en ce qui concerne l'hypergamie.

1 Littérature : des débats centrés sur les effets du développement et de l'ouverture sociale

Comme pour l'étude des évolutions temporelles, la littérature existante concernant les variations entre pays et leurs déterminants concerne principalement l'homogamie d'éducation. À notre connaissance, aucune analyse des déterminants de l'hypergamie n'a été publiée à ce jour, à l'exception de l'étude d'Albert Esteve, Joan García-Román et Iñaki Permanyer (2012), qui mesurait l'effet de la supériorité féminine en termes d'éducation sur la composition des couples (*cf.* chapitre 4).

Les comparaisons internationales concernant l'intensité de l'homogamie, plus nombreuses et détaillées, remontent à la contribution de John Hall (1954) dans l'ouvrage collectif *Social Mobility in Britain*. Mettant en regard dans une courte note l'Angleterre, le Pays de Galles et les États-Unis, l'auteur concluait que l'homogamie d'origine sociale (profession du père en trois catégories) était d'ampleur comparable dans les trois sociétés. Homogamie et mobilité sociale étant souvent considérées comme liées depuis les travaux de Pitirim Sorokin ([1927] 1998), cette première observation a contribué à la conclusion de Seymour Martin Lipset et Hans Zetterberg (1959) selon laquelle la structure de la mobilité sociale intergénérationnelle et de l'hétérogamie est très largement similaire parmi les pays occidentaux. Comme dans le cas de la mobilité sociale, cependant, cette hypothèse a été depuis dépassée par des analyses plus détaillées. David Featherman, Frank Lancaster Jones et Robert Hauser (1975) ont avancé l'hypothèse plus restreinte selon laquelle les niveaux de mobilité relative (une fois contrôlée la structure de la population) entre pays à économie de marché et à système familial nucléaire sont similaires, une fois un certain niveau de développement atteint. Transposée à l'homogamie, cette hypothèse moins forte n'apparaît pas non plus vérifiée à strictement parler, même si elle constitue comme on le verra une approximation raisonnable en termes de description des couples.

Les premiers à avoir remis en cause l'hypothèse Lipset-Zetterberg, Wout Ultee et Ruud Luijkx (1990), se sont fondés sur des données portant sur 23 pays industrialisés depuis 1945, pour mettre en évidence des variations notables dans l'intensité de l'homogamie relative et

de la mobilité sociale intergénérationnelle aussi bien entre pays qu'au cours du temps. Ils notaient en revanche qu'homogamie et hérédité sociale allaient de pair, confortant l'interprétation de S. M. Lipset et H. Zetterberg selon laquelle ces deux phénomènes sont liés, et peuvent être considérés comme des indices de l'ouverture générale d'une société. Les auteurs ont par ailleurs montré que le niveau de développement (mesuré au nombre de lignes téléphoniques par habitant), la faible ségrégation scolaire (approchée par l'absence de paliers d'orientation avant 12 ans) et une longue présence de partis sociaux-démocrates au gouvernement sont associés à une homogamie relative plus faible. Contrairement à ce que les auteurs attendaient, les pays socialistes présentent une homogamie plus forte que les pays capitalistes une fois contrôlées les autres variables.

1.1 L'hypothèse d'une relation en U inversé avec le développement

Cette veine de recherche a été poursuivie dans une série d'articles par un collectif de chercheurs autour de Jeroen Smits. Une première étude (Smits, Ultee et Lammers, 1998) a analysé les déterminants de l'homogamie relative d'éducation dans 65 pays de manière transversale, à partir de données portant sur les années 1970. Les auteurs ont observé une relation en U inversé entre niveau de développement (mesuré à la consommation d'énergie par habitant et à la proportion de la population active non agricole) et homogamie : en cohérence avec le cadre théorique qu'ils ont élaboré, l'homogamie croît dans la première phase de l'industrialisation, puis décroît ensuite. Elle est donc la plus forte parmi les pays de niveau de développement intermédiaire. Cette évolution est interprétée comme la conséquence de l'importance croissante prise par l'éducation dans un premier temps, puis par l'individualisation des sociétés et l'augmentation des contacts entre groupes dans un second temps.

L'étude a aussi mis en évidence une moindre tendance à l'homogamie dans les pays dont la religion dominante est le protestantisme par rapport aux pays à dominante catholique, à la fois catholique et protestante, musulmane ou confucéenne, ce que les auteurs attribuent à une moindre importance accordée aux valeurs familiales. Il paraît cependant difficile d'exclure que cette différence soit due à d'autres particularités des pays protestants de l'échantillon considéré (pays scandinaves et anciennes colonies britanniques, au nombre de 7). Finalement, les pays en développement dont l'agriculture avait atteint le

stade agrarien (système fondé sur le labour, selon la distinction de Lenski et Nolan, 1984), aux inégalités plus marquées, présentent une homogamie plus élevée que ceux qui partaient du stade de l'horticulture.

La conclusion de cet article concernant la plus forte homogamie des pays de tradition confucéenne a été critiquée par James M. Raymo et Yu Xie (2000), qui ont au contraire observé des niveaux d'homogamie très similaires entre les États-Unis et le Japon, et un niveau encore inférieur à Taïwan. Cependant, l'hypothèse d'une évolution en U inversé a trouvé une confirmation partielle dans ce nouvel échantillon plus réduit de pays. Surtout, les auteurs ont souligné le saut interprétatif qui consiste à inférer une tendance temporelle à partir d'une comparaison transversale entre sociétés : tous les pays ne suivent pas nécessairement la même trajectoire. La méthode consistant à « lire l'histoire de côté » (Thornton, 2001) a déjà conduit les sciences sociales à prendre des différences durables entre régions du monde pour les signes d'une évolution inéluctable vers un état considéré comme moderne.

Répondant à cette critique, les mêmes auteurs (Smits, Ultee et Lammers, 2000) ont enrichi leur précédente analyse en distinguant, au sein de chaque pays, deux cohortes de naissance. Ils en ont conclu que les pays dans lesquels l'homogamie est plus faible de manière statistiquement significative parmi les cohortes les plus récentes sont plus nombreux que les cas inverses (15 contre 7) – résultat interprété comme le signe d'une baisse de l'homogamie au cours du temps. La baisse apparaît plus forte dans les pays les moins développés, et dans ceux au développement le plus rapide, mais elle ralentit, voire s'inverse au-delà d'un certain niveau. Cette tendance temporelle soutient donc l'hypothèse d'une certaine saturation de la croissance de l'hétérogamie, et donc d'une convergence, parmi les pays les plus développés, quelles que soient leurs différences culturelles et religieuses. Néanmoins, comme nous l'avons déjà souligné au chapitre 1, la comparaison de cohortes de naissance à partir d'une seule enquête transversale n'est pas une méthode très fiable pour mesurer les tendances temporelles. Une certaine prudence nous semble donc de mise dans l'interprétation de ces résultats.

Comme prolongement des études précédentes, J. Smits (2003) a concentré son attention sur l'endogamie relative des diplômés du secondaire et du supérieur dans 55 pays. Observant comme précédemment que l'endogamie de ce groupe décroît avec le niveau de

développement, et que cette baisse est d'autant plus rapide que la croissance est élevée et que le niveau de développement est faible, il a aussi confirmé qu'elle est plus faible dans les pays à dominante protestante. Fait nouveau, l'expansion du groupe apparaissait comme négativement associée à son endogamie, ce que l'auteur expliquait par deux facteurs : tout d'abord, les petits groupes sont dotés d'une plus forte conscience de leur situation ; ensuite, la valeur sociale des diplômés est d'autant plus élevée, et donc le groupe des diplômés spécifique, qu'ils sont rares.

Enfin, concluant cette série de travaux, J. Smits et Hyunjoon Park (2009) ont analysé les variations de l'homogamie d'éducation dans dix pays asiatiques, aussi bien entre pays qu'entre cohortes de mariages. Confirmant les résultats précédents, ils ont observé une diminution temporelle de l'homogamie relative au fil des cohortes de mariage dans tous les pays étudiés, dans un contexte de développement économique et de massification scolaire rapides. Le degré de modernisation, la proportion de diplômés du supérieur et le taux d'activité des femmes apparaissaient associés à une faible homogamie, alors que l'influence confucéenne était corrélée à une plus forte homogamie.

Il semble nécessaire de souligner certaines limites de ces travaux. Il s'agit premièrement de la qualité parfois faible des jeux de données, soulignée par les auteurs eux-mêmes : l'homogamie est mesurée à partir de sources hétérogènes, et les classifications ne sont pas toujours très bien harmonisées ; les variables explicatives sont souvent frustes du fait de la non disponibilité des informations adéquates pour l'ensemble des pays. La force de ces études, tenant au large échantillon de pays étudiés, en constitue aussi la principale faiblesse – critique qui s'applique moins à la dernière étude citée (Smits et Park, 2009). Ces limites concernent deuxièmement le champ d'analyse, restreint aux couples mariés, ce qui peut conduire à surestimer la force de l'homogamie dans les pays où la cohabitation est la plus développée – comme nous l'avons vu au chapitre 1.

1.2 Les études récentes : des données harmonisées

Les progrès récents des enquêtes quantitatives internationales et de l'harmonisation des enquêtes nationales ont rendu possibles des études à partir de bases de données plus homogènes. En se fondant sur l'Enquête sociale européenne (*European Social Survey*) 2004-2005, Henryk Domański et Dariusz Przybysz (2007) ont étudié les variations de l'homogamie

d'éducation dans 22 pays européens. Cette étude sur un champ géographique plus restreint que les précédentes a observé que l'homogamie d'éducation diminuait d'abord avec le développement (mesuré au produit intérieur brut par habitant), avant de remonter aux niveaux les plus avancés, contredisant l'hypothèse de l'évolution en U inversé pour les pays les plus développés. Elle a aussi mis en évidence des différences entre groupes de pays : l'homogamie apparaît la plus forte dans les pays de l'ancien bloc soviétique, et la plus faible parmi les pays à régime social-démocrate. Concernant la religion dominante, l'homogamie apparaît plus forte dans les pays protestants. Ce dernier résultat, en contradiction avec les études précédentes qui ne contrôlaient pas le régime d'État-providence, illustre à nos yeux la difficulté à séparer clairement des dimensions qui sont inextricablement liées dans la réalité des configurations nationales.

Malheureusement, l'étude ne portait que sur les couples mariés, alors que les auteurs relevaient des variations importantes du taux de cohabitation hors mariage (jusqu'à 20 % en Suède). En outre, l'analyse des déterminants de l'homogamie relative portait exclusivement sur l'intensité de l'association parmi les couples hétérogames¹, ignorant de ce fait l'une des dimensions principales du phénomène. Enfin, l'existence d'une relation en U entre niveau de développement et homogamie, estimée à partir d'un modèle de régression linéaire comprenant le produit intérieur brut par habitant et son carré, n'était accompagnée d'aucun détail permettant d'évaluer la qualité d'ajustement et la signification de cette courbe – ce qui est généralement le cas dans les travaux de ce type. Pourtant, les pays à la plus forte homogamie (Slovaquie, République tchèque, Norvège, Allemagne, Ukraine, Pologne, Hongrie, Estonie, Slovénie et Grèce) n'apparaissent pas véritablement mieux placés en termes de développement que les pays à l'homogamie la plus faible (Luxembourg, France, Suède, Finlande, Belgique et Espagne). Le fait que les auteurs aient contrôlé simultanément la religion dominante et le régime d'État-providence brouille passablement l'interprétation de ce résultat.

Toujours à partir de l'Enquête sociale européenne, mais en retenant les vagues 2002, 2004 et 2006, Tomáš Katrňák, Petr Fučík et Ruud Luijkx (2012) ont étudié le lien entre homogamie d'éducation, mobilité intergénérationnelle en termes de diplôme et inégalités scolaires dans 29 pays européens. Ils concluent à une forte corrélation (entre 0,8 et 0,9) entre

1 Précisément, la modélisation portait sur le paramètre de pas d'un modèle de distance comprenant des coefficients diagonaux spécifiques à chaque pays.

intensité de l'homogamie d'éducation et hérédité éducative en termes relatifs (soit l'inverse de la mobilité sociale relative), signe supplémentaire que les deux dimensions peuvent être considérées comme l'indice d'un même phénomène, l'ouverture sociale. Cependant, l'homogamie est systématiquement plus forte que l'immobilité intergénérationnelle, et constitue donc pour les auteurs un meilleur indicateur du degré d'ouverture d'une société du point de vue de l'éducation. L'étude a enfin confirmé les résultats de la précédente (Domański et Przybysz, 2007) selon laquelle l'homogamie relative d'éducation est la plus forte parmi les pays de l'ancien bloc soviétique, et la plus faible parmi les pays scandinaves, en France et en Grande-Bretagne². Notons qu'encore une fois, seuls les couples mariés étaient pris en compte.

Enfin, une étude récente à partir de la Luxembourg Income Study (Monaghan, 2015) a porté sur le lien entre homogamie d'éducation et inégalités de revenus dans 25 pays développés, à partir de plusieurs enquêtes par pays depuis 1980. L'auteur a observé un lien positif entre homogamie relative d'éducation et rendements de l'éducation en termes de revenu – autrement dit, inégalités de revenus entre groupes de diplôme. Il a aussi mesuré une corrélation positive entre écart de taux d'activité des hommes et des femmes et homogamie d'éducation, confirmant ainsi que l'augmentation de l'activité professionnelle féminine est associée à une moindre homogamie.

1.3 Une synthèse possible ?

Malgré des divergences importantes, on peut tenter de résumer ces différentes études en quelques faits saillants. L'hypothèse selon laquelle l'homogamie et l'hypergamie ne varient pas entre pays développés est clairement rejetée. Comme pour la mobilité sociale intergénérationnelle, avec laquelle elle apparaît fortement liée, des différences importantes ont pu être relevées même une fois contrôlée la structure de la population. L'hypothèse Featherman-Jones-Hauser selon laquelle ces différences devraient disparaître au-delà d'un certain niveau de développement ne paraît pas non plus soutenue par les études portant sur l'Europe, qui mettent par exemple en évidence des différences entre Allemagne, Italie et France.

2 Certaines différences notables peuvent cependant être relevées, comme la très faible homogamie de l'Estonie et de la Norvège, qui figuraient au contraire parmi les pays les plus homogames dans l'étude précédente.

Il reste cependant à déterminer quel lien entretient l'homogamie avec les différents stades du développement, et en particulier si l'hypothèse d'une relation en U inversé est vérifiée. L'étude de H. Domański et D. Przybysz (2007), qui conclut à un renforcement de l'homogamie aux niveaux de développement les plus élevés, nous semble être l'exception, et demander une confirmation. Quelle que soit la validité de l'hypothèse d'une évolution en U inversé, il nous semble que les preuves empiriques tendent plutôt dans le sens d'un affaiblissement de l'homogamie aux stades les plus avancés du développement. Corroborant cette conjecture, on peut relever qu'au niveau européen, les anciens pays communistes, moins développés que leurs voisins d'Europe de l'Ouest, présentent une plus forte homogamie d'éducation. Il n'est pas possible de mettre à l'épreuve la première partie de la relation en U inversé, qui concerne le passage d'un niveau de développement faible à un niveau intermédiaire. En revanche, le cas européen est adapté à l'évaluation de l'hypothèse d'une saturation, ou du moins d'un ralentissement de la baisse de l'homogamie, aux niveaux de développement les plus élevés.

Concernant la question de l'ouverture sociale, il semble bien que mobilité sociale et hétérogamie aillent de pair, et mesurent ainsi en grande partie le même phénomène social. De même, les quelques résultats disponibles incitent à penser qu'homogamie d'éducation et inégalités de revenus sont positivement corrélées. Si la première conclusion est cohérente avec les résultats des études portant sur la mobilité sociale, la seconde l'est moins. En effet, dans le cas de l'Europe, l'ouvrage collectif dirigé par Richard Breen (2004, p. 396) n'a pas réussi à identifier de relation claire entre mobilité sociale relative et inégalités. Il n'est donc pas certain que cette relation, identifiée en comparant des pays développés assez divers, soit aussi valable à l'intérieur de l'Europe.

L'effet de l'augmentation de l'activité féminine semble clair. Contrairement à ce que certaines conjectures pouvaient laisser penser, cette évolution n'entraîne pas de renforcement de l'homogamie, qui découlerait d'une plus grande préférence pour la similarité des hommes et des femmes.

Finalement, l'influence de la religion n'a pas été clairement établie. On peut penser que la forte corrélation entre la religion dominante d'un pays et toute une série de caractéristiques historiques rend très difficile, voire dépourvue de sens, l'étude de l'influence

de cette dimension sur l'homogamie à caractéristiques contrôlées : selon les variables incluses dans le modèle, les résultats peuvent varier très fortement.

Dans la suite de ce chapitre, nous confirmerons l'importance des variations de l'homogamie et de l'hypergamie entre pays, et montrerons que ce constat s'applique aussi aux régions à l'intérieur de ceux-ci. Les déterminants de ces variations seront évoqués lorsqu'une description qualitative s'avère possible, avant un examen plus systématique au chapitre suivant.

2 Données : l'Enquête européenne sur les forces de travail

2.1 Choix et définition des unités géographiques : les régions NUTS1

L'étude est réalisée à partir des données de l'Enquête européenne sur les forces de travail (*European Union Labour Force Survey*, EU-LFS) 2011, qui résulte de la mise en commun par l'institut statistique européen Eurostat des enquêtes Emploi conduites au niveau national par chacun des États membres³. 23 pays de l'Union européenne, divisés en 64 régions, ont pu être retenus : Allemagne (DE), Autriche (AT), Belgique (BE), Bulgarie (BG), Chypre (CY), Espagne (ES), Estonie (EE), Irlande (IE), Italie (IT), France (FR), Grèce (GR), Hongrie (HU), Lettonie (LV), Lituanie (LT), Luxembourg (LU), Pays-Bas (NL), Pologne (PO), Portugal (PT), République Tchèque (CH), Roumanie (RO), Royaume-Uni (UK), Slovaquie (SK) et Slovénie (SI).

Les couples cohabitants, mariés ou non, sont identifiés au sein de chaque ménage à partir des identifiants des conjoints (voir ci-dessous). Cette méthode ne permet pas de reconstruire les couples dans les enquêtes nationales qui ne sont pas conduites au niveau des ménages : pour cette raison, les pays scandinaves (Danemark, Finlande, Norvège⁴ et Suède), dont les données proviennent de registres de population (voir par exemple Villund, 2010 concernant la Norvège) n'ont pu être retenus dans l'échantillon. Cette exclusion est clairement préjudiciable à notre analyse, dans la mesure où les études de la stratification sociale ont mis en évidence la structure nettement plus égalitaire de ces sociétés par rapport

3 Ces données nous ont été mises à disposition par Eurostat dans le cadre du projet d'élaboration d'une classification socioéconomique européenne ESeG (*European Socio-Economic Groups*).

4 La Norvège, bien que non membre de l'Union européenne, participe à l'Enquête européenne sur les forces de travail à titre de pays associé.

au reste de l'Europe (Domański et Przybysz, 2007 ; Katrňák, Fučík et Luijkx, 2012). Néanmoins, les pays considérés demeurent suffisamment divers pour permettre l'observation de différences importantes et l'analyse des facteurs qui en sont à l'origine.

La définition de notre échantillon s'éloigne ainsi nettement d'un tirage aléatoire de pays qui nous permettrait de prétendre à une hypothétique représentativité par rapport à une super-population qui correspondrait par exemple à l'ensemble des pays du monde. Cette situation est classique dans le cadre d'analyses comparatives, qui sont généralement dépendantes de facteurs institutionnels – et donc historiques – qui affectent la disponibilité ou non de données (Ebbinghaus, 2005). Elle ne pose pas de problème particulier, dans la mesure où nous ne recourons à l'inférence statistique que pour passer de notre échantillon d'individus à la population totale des régions considérées : nous ne faisons jamais l'hypothèse d'un échantillonnage aléatoire par tirage de régions. En cohérence avec cette approche, nous pondérons toutes nos analyses relativement à la population de chaque région, considérant en cela que l'étude vise à décrire la population de l'Union européenne, plutôt que des pays quelconques considérés chacun comme une unité d'observation indépendante d'importance égale à toutes les autres (Ebbinghaus, 2005, p. 136).

L'échelle d'analyse élémentaire est celle de la région NUTS1 (niveau supérieur de la Nomenclature des unités territoriales pour les statistiques). Ces régions, qui regroupent entre 3 et 7 millions de personnes, correspondent aux Länder allemands, aux zones économiques d'aménagement du territoire françaises (regroupant différentes Régions), ou encore à des regroupements de communautés autonomes espagnoles. Pour des raisons de taille d'échantillon, certaines régions ont dû être fusionnées. C'est le cas des Länder allemands, qui, de 16 à l'origine, sont regroupés en trois régions : Est (dont Berlin), Nord-Ouest, et Sud-Ouest. Ainsi que des régions NUTS1 du Royaume-Uni, qui passent de 12 à 6 : Nord de l'Angleterre ; Midlands ; Sud et Est de l'Angleterre ; Londres ; Pays de Galles et Irlande du Nord ; Écosse. Ces regroupements ont été réalisés sur la base d'une continuité géographique, en tenant autant que possible compte de leur histoire. Les informations concernant les régions ne sont pas disponibles dans le cas des Pays-Bas, qui sont donc considérés comme une seule région. Enfin, pour des raisons d'homogénéité géographique et culturelle, ainsi que du fait de la faible taille des échantillons correspondants, les régions d'outre-mer ont été exclues : Canaries (ES7), départements français d'outre-mer (FR9),

Açores (PT2) et Madère (PT3). La liste complète des régions retenues, de leurs noms et de leurs codes est détaillée en appendice dans le tableau 7.5 (p. 454).

2.2 Variables : niveau d'éducation et groupe socioéconomique

Deux composantes de l'homogamie et de l'hypergamie sont considérées : l'éducation, et le groupe socioéconomique. Aucune information concernant l'origine sociale n'étant disponible dans l'Enquête européenne sur les forces de travail, nous ne pourrions traiter de cette dimension ici.

2.2.1 Le niveau d'éducation

Le niveau d'éducation⁵ des conjoints est mesuré de manière malheureusement relativement peu détaillée. En effet, en dépit des efforts d'harmonisation conduits par l'Unesco, l'OCDE et Eurostat pour la définition et la mise en œuvre de la Classification internationale type de l'éducation (dite dans la version utilisée CITE-97, ou ISCED-97 en anglais), la codification des diplômes nationaux reste très imparfaite dans l'Enquête européenne sur les forces de travail. Ainsi, alors que les données fournies par Eurostat distinguent potentiellement 13 catégories, reprenant et affinant la CITE, nombre d'entre elles ne sont utilisées que par une poignée de pays (Schneider et Kogan, 2008, p. 34-42 ; Schneider, 2009, sect. 4.1.3). Une analyse portant sur l'ensemble des pays doit *in fine* se limiter à la classification en trois niveaux qui est construite par Eurostat (Tableau 7.1) pour ne pas faire face à trop de catégories vides et ne pas trop s'exposer à des incohérences criantes de classement. Trois distinctions importantes sont en particuliers impossibles à réaliser de manière correcte dans tous les pays.

5 Contrairement aux analyses portant sur la France, il ne s'agit pas ici à proprement parler d'un diplôme étant donné que la nomenclature utilisée tient compte du niveau des études les plus élevées suivies, et pas uniquement de leur certification.

Tableau 7.1 : La classification d'éducation en trois niveaux de l'Enquête européenne sur les forces de travail

Intitulé	Catégories CITE-97	Correspondance avec le système français
≤ Scolarité obligatoire	Pré-primaire (0)	Aucun diplôme
	Primaire (1)	CEP
	1 ^{er} cycle du secondaire (2)	BEPC
	2 nd cycle du secondaire professionnalisant court (3C-court)	–
Secondaire complet	2 nd cycle du secondaire professionnalisant long (3C-long)	CAP, BEP
	2 nd cycle du secondaire donnant accès aux études supérieures (3AB)	Bac général, technologique et professionnel
	Post-secondaire (4)	Capacité en droit
Supérieur	Supérieur professionnel (5B)	Licence professionnelle, BTS, DUT
	Supérieur général (5A)	DEUG, DEUST, Licence, Master, Grandes écoles
	Qualification pour la recherche (6)	Doctorat

Les équivalences avec le système français, fournies pour illustrer ces catégories et pour faciliter la comparaison avec nos analyses précédentes portant sur la France, sont fondées sur les descriptions fournies par Annick Kieffer (2008).

Premièrement, la catégorie des personnes ayant quitté le système scolaire sans valider aucun diplôme ne peut être définie de manière cohérente à travers l'Europe (Schneider, 2008a, p. 33, 2009, p. 164). Dans de nombreux pays (par exemple le Royaume-Uni, *cf.* Schneider, 2008b, p. 293), tous les individus scolarisés jusqu'à l'âge minimum obligatoire sont classés au niveau du premier cycle du secondaire (niveau 2 de la CITE) : la proportion de non diplômés est résiduelle, voire parfois nulle parmi les générations récentes. Cette convention diffère nettement de celle retenue par exemple en France (Kieffer, 2008, p. 116), où les individus n'ayant pas validé leur BEPC ou leur Brevet des collèges sont classés au niveau primaire (1) s'ils n'ont pas prolongé leur scolarité au-delà du collège. Par conséquent, retenir une distinction en dessous du premier cycle du secondaire, loin d'améliorer la précision des tables d'homogamie, ne ferait qu'introduire des artefacts selon les décisions de recodage prises par chaque pays.

Deuxièmement, il n'est pas possible de distinguer entre différentes filières au niveau du second cycle du secondaire, catégorie hétérogène qui constitue le groupe le plus nombreux

parmi les cohortes récentes. Les 13 catégories retenues par l'Enquête européenne sur les forces de travail ne séparent ainsi pas les filières ouvrant la voie vers des études supérieures générales (3A) de celles menant à des études supérieures professionnelles (3B), alors que cette distinction est essentielle dans de nombreux systèmes scolaires (Schneider et Kogan, 2008, p. 36). En outre, la distinction entre filières menant à des études supérieures professionnelles (3B) et celles préparant à une entrée directe dans la vie professionnelle (3C), même si elle est théoriquement présente dans les données, est souvent faite de manière incohérente. Sans que les caractéristiques des cursus concernés justifient nettement ces décisions, la filière 3C est ainsi réservée par de nombreux pays à des cursus marginaux en termes d'effectifs, qui classent en 3B la quasi-totalité de leurs filières professionnelles du secondaire. Ainsi, en Allemagne, les diplômes professionnels du second cycle du secondaire (apprentissage) sont classés officiellement dans la CITE comme donnant accès à des études supérieures professionnelles (3B), plutôt que comme relevant de filières directement professionnalisantes (3C). Cette dernière option est pourtant la plus correcte (Schneider, 2008c, p. 94-95) dans la mesure où ces diplômes préparent à l'entrée directe sur le marché du travail, et ne donnent accès à des études supérieures professionnelles qu'après quelques années de vie active. De même, en Espagne (Ortiz, 2008, p. 132-146), des filières qui relèveraient du secondaire professionnalisant (3C) sont considérées de manière abusive comme menant à des études supérieures professionnelles (3B), voire comme des formations du supérieur professionnel (5B).

D'autres pays ne fournissent pas des informations assez détaillées pour distinguer ces filières. En Slovénie, pays dans lequel une filière préparant aux études supérieures (3AB) et une filière professionnalisante (3C) aux effectifs comparables existent, les deux formations sont confondues dans une seule catégorie du fait d'un manque de détail de la variable collectée dans l'enquête nationale (Ivančič, 2008, p. 191). Sans que l'on dispose d'une analyse aussi détaillée, il semble que le Portugal soit dans une situation similaire, étant donné que la distinction entre 3C et 3AB, pourtant présente dans la documentation officielle de la CITE, n'est pas reprise dans l'Enquête européenne sur les forces de travail. Au total, pour corriger ces divergences, la classification en trois niveaux construite par Eurostat rattache les filières les plus courtes (moins de deux ans) du secondaire professionnalisant (3C) au premier cycle du secondaire, et les plus longues au second cycle (Tableau 7.1).

Troisièmement, des difficultés similaires apparaissent lorsque l'on tente de distinguer entre filières professionnelles et générales du supérieur. En Italie, des formations qui relèveraient plutôt du professionnel ont été considérées comme générales (Barone et Schizzerotto, 2008, p. 154), retirant à la catégorie une grande partie de ses effectifs. En Hongrie, les formations dispensées par les hautes écoles sont considérées comme des filières générales alors qu'elles relèvent plutôt du professionnel (Bukodi, Róbert et Altorjai, 2008, p. 208) : ici encore, les effectifs du supérieur professionnel apparaissent par conséquent presque nuls. En Pologne, les données de l'Enquête européenne sur les forces de travail ne distinguent pas filières générale et professionnelle du supérieur, alors que cette distinction existe dans les enquêtes nationales (Mach et Kryszczuk, 2008, p. 234). Nous sommes donc forcés de confondre l'ensemble des diplômés du supérieur dans une même catégorie⁶.

Ayant souligné les limites du codage de l'éducation dans l'Enquête européenne sur les forces de travail, il nous faut cependant relativiser l'incidence de ce problème sur la mesure des variations de l'intensité de l'homogamie entre régions d'Europe. Si le manque de détail interdit toute description fine des différences dans la structure de l'homogamie – comme un potentiel déplacement vers le haut des barrières sociales avec la massification scolaire –, son niveau global apparaît relativement robuste à des changements de nomenclature. Conformément à ce que notaient déjà J. Smits, W. Ultee et J. Lammers (1998, n. 4), nous observons que la corrélation entre les indices d'intensité de l'homogamie relative des différents pays ou régions obtenus à partir de classifications plus ou moins détaillées est supérieure à 0,9. Ce résultat semble contredire les remarques méthodologiques formulées au chapitre 1, où nous soulignons que le recours à des classifications trop agrégées pouvait masquer des tendances visibles seulement à un niveau de détail élevé (Wong, 2003). Mais, comme on le verra plus bas, les différences entre régions européennes sont bien plus marquées que les évolutions temporelles à l'intérieur d'un même pays : l'imperfection de l'instrument de mesure porte ici moins à conséquence. Il reste que l'on ne peut exclure qu'une classification plus détaillée permettrait d'affiner nos résultats en mettant en évidence des différences entre régions qui apparaissent ici proches en termes d'intensité de l'homogamie d'éducation.

6 La CITE-97 ne permet pas non plus de distinguer les diplômés du supérieur selon la durée du cursus, à l'exception du doctorat (niveau 6), qui correspond à des effectifs trop faibles pour justifier la création d'une catégorie. Cette limitation a été corrigée dans la nouvelle version de la classification, publiée en 2011.

Pour finir, notons néanmoins que le caractère fruste de la classification retenue ne doit pas occulter la qualité relativement élevée des données et du codage réalisé par Eurostat, une fois le niveau de détail limité à trois niveaux. Les autres sources de données disponibles, dont notamment l'Enquête sociale européenne (*European Social Survey*, ESS), ayant connu les mêmes difficultés d'harmonisation (Schneider, 2009, p. 90), n'offrent pas des informations plus fiables. En attendant une meilleure standardisation des classifications de niveau d'éducation, ces enquêtes constituent cependant un clair progrès par rapport à l'analyse de tables d'homogamie collectées dans les classifications nationales, que les chercheurs n'ont pas toujours individuellement les moyens d'harmoniser correctement.

2.2.2 Le groupe socioéconomique

L'identification du groupe socioéconomique des conjoints soulève moins de difficultés. Celui-ci est mesuré dans le premier niveau de la récente classification ESeG (*European Socio-Economic Groups*), en sept catégories d'actifs (Encadré 7.1). Cette nomenclature a été retenue, de préférence notamment à la classe EGP utilisée pour la France, pour plusieurs raisons. La première est institutionnelle : la présente étude ayant été menée dans le cadre du projet ESeG, il était logique de mobiliser la classification construite par le projet, contribuant par là à l'illustration de son utilité et à sa diffusion dans la sociologie européenne. La seconde concerne le codage de la variable à partir de l'Enquête européenne sur les forces de travail 2011 : cette vague d'enquête étant la première à coder la profession dans la Classification internationale type des professions 2008 (CITP-08, ISCO-08 en anglais), aucune table n'existe encore pour coder la classe EGP à partir de ces informations, alors que le projet ESeG s'est précisément fondé sur cette version de la CITP. Enfin, une dernière raison tient à la taille des échantillons disponibles pour chaque région : plus petits que ceux utilisés pour décrire l'ensemble de la France, ceux-ci ne permettraient pas de retenir une classification aussi détaillée que la version d'EGP en 11 catégories mobilisée sur les données françaises.

L'absence d'un groupe séparé pour les agriculteurs, qui sont confondus avec les autres petits indépendants dans le niveau 1 de la nomenclature ESeG, soulève une question délicate. Il pourrait en effet sembler nécessaire de distinguer ce groupe des autres indépendants, du fait de sa très forte endogamie, et de sa place particulière d'un point de vue théorique : le déclin numérique des agriculteurs et l'affaiblissement de leur endogamie

relative sont à l'origine d'une part importante de la diminution de l'endogamie en France au cours des quarante dernières années, et donc potentiellement aussi source de différences entre régions européennes de niveaux de développement inégaux. En effet, les régions les plus rurales de Roumanie, de Grèce, de Pologne et du Portugal comprennent entre 10 et 40 % d'agriculteurs au sein de leur population active (voir colonne « Agriculture » dans le tableau 8.3, p. 496). Cependant, cette position se heurte à une difficulté pratique : dans la moitié des régions, les agriculteurs représentent moins de 2 % des actifs des deux sexes. De ce fait, ce groupe est complètement absent des échantillons de nombreuses régions ; 31 d'entre eux contiennent moins de 10 conjoints agriculteurs ou de 10 conjointes agricultrices parmi les couples retenus. Ainsi, la mesure de l'homogamie parmi les agriculteurs est impossible dans certains cas, et très imprécise dans la plupart des autres. En outre, dans les pays dans lesquels le groupe conserve une taille importante, la structure de l'association est presque identique chez les petits indépendants et chez les agriculteurs : même dans ces cas, la distinction n'apporte que peu à l'analyse. Nous avons donc décidé de suivre le choix réalisé lors de l'élaboration de la nomenclature ESeG en ne distinguant pas les agriculteurs des autres petits indépendants.

Afin de limiter l'influence des différences de taux d'activité féminine entre régions, le groupe socioéconomique des inactifs est défini à partir de leur dernière profession. Cette information n'est disponible que lorsque la personne a occupé un emploi dans les huit années précédant l'enquête⁷. Nous ne présentons pas les résultats calculés à partir des seuls couples bi-actifs, qui sont quasiment identiques à ceux obtenus pour l'ensemble des couples (corrélations des variations de l'intensité de l'homogamie relative supérieure à 0,9). Il semble qu'au sein d'une région donnée, l'association entre professions ou anciennes professions des conjoints diffère peu selon le statut d'activité – du moins au regard de l'ampleur des variations entre régions.

7 Le dernier emploi des inactifs n'est pas disponible dans les données françaises avant 2012, le codage de la CITP-08 n'ayant pas été réalisé pour ces individus. Cette limitation n'est pas un véritable problème dans la mesure où la prise en compte des inactifs ne fait presque aucune différence pour ce pays dans les années récentes (chapitre 2), et en particulier parmi les jeunes couples.

Encadré 7.1 : La classification socioéconomique européenne ESeG

La nomenclature des Groupes socioéconomiques européens (*European Socio-Economic Groups*, ESeG) a été élaborée entre 2011 et 2014 par un groupe de travail international de statisticiens et de chercheurs piloté par l'Insee (Meron et Amar, 2014). Ce projet faisait suite à un appel d'offres de l'institut statistique européen Eurostat en vue de proposer une alternative à la classification ESeC qui avait fait l'objet d'un précédent appel d'offres suivi d'une élaboration entre 2004 et 2006 (Rose et Harrison, 2009). L'objectif de cette nouvelle classification était de pouvoir être directement mise en œuvre dans les enquêtes du système statistique européen, et notamment dans l'Enquête européenne sur les forces de travail. Pour cela, adoptant une approche plus pragmatique que son prédécesseur ESeC, la classification ESeG se fonde sur des variables essentielles disponibles dans les enquêtes à partir de 2011 : profession exprimée dans le niveau 2 de la nomenclature internationale type des professions (CITP-08), statut d'emploi (indépendant ou salarié), secteur d'activité.

Plusieurs prototypes de nomenclature ont été mis à l'épreuve à partir notamment des données de l'Enquête européenne sur les forces de travail et celles des Statistiques sur les revenus et les conditions de vie (SRCV/EU-SILC). Notre participation au projet a consisté à évaluer ces prototypes au regard de l'homogamie socioéconomique, ainsi qu'à évaluer la pertinence des regroupements possibles pour constituer le niveau 2 de la classification (Bouchet-Valat, 2014a, 2014b).

Le niveau 1 de la nomenclature, en sept postes d'actifs et deux postes pour les inactifs, ne recourt qu'au premier niveau de la classification CITP, de manière à permettre un codage fiable dans un grand nombre d'enquêtes. Il est défini comme suit :

- Cadres (salariés mais aussi cadres dirigeants de grandes entreprises)
- Professions libérales et intellectuelles
- Professions intermédiaires
- Petits indépendants (avec ou sans salariés, dont les agriculteurs exploitants)
- Employés qualifiés
- Ouvriers qualifiés

- Professions peu qualifiées (dont les ouvriers agricoles)
- Inactifs de moins de 65 ans
- Retraités et inactifs de plus de 65 ans

On voit que cette classification est relativement similaire à la nomenclature EGP ou à celle des PCS. Par rapport à EGP, les cadres et professions intellectuelles supérieures se voient divisées en deux groupes, alors qu'à l'inverse les ouvriers et employés non qualifiés sont confondus dans un seul groupe, celui des professions peu qualifiées. Par rapport aux PCS, les employés et les ouvriers sont séparés selon leur qualification, et les dirigeants de grandes entreprises sont classés avec les cadres. Le premier niveau d'ESeG diffère de ces deux classifications par le fait que les agriculteurs sont confondus avec les artisans et commerçants dans un seul groupe d'indépendants, du fait de la très faible représentation de ce groupe dans plusieurs pays européens ; ils sont cependant distingués dès le niveau 2 de la classification. Enfin, les frontières des différents groupes ne correspondent pas toujours entre les trois classifications, notamment en ce qui concerne les professions intermédiaires.

2.2.3 La définition de l'hypergamie

L'hypergamie féminine d'éducation est simplement définie comme le fait que l'homme possède un niveau d'éducation supérieur à celui de sa conjointe dans la classification en trois catégories présentée ci-dessus.

La mesure de l'hypergamie socioéconomique suppose de se doter d'un ordre hiérarchique des groupes socioéconomiques, défini suivant la matrice présentée dans le tableau 7.2. Cet ordre est construit selon les mêmes principes qu'au chapitre 4 (p. 251). Les catégories salariées de la classification ESeG sont assez nettement hiérarchisées, à l'exception d'une part des cadres et des professions libérales et intellectuelles, et de l'autre des employés et ouvriers qualifiés, que nous considérons ici respectivement comme de même niveau. Les petits indépendants, qui incluent ici les agriculteurs à la différence de la classification EGP, sont considérés comme formant des couples non hiérarchisés, sauf

lorsque leur conjoint occupe une profession peu qualifiée, statut considéré comme inférieur à l'indépendance.

Tableau 7.2 : Matrice d'hypergamie socioéconomique

Hommes	Femmes				
	CPLI	PI	E&OQ	PPQ	I
Cadres et professions libérales et intellectuelles	H=F				
Professions intermédiaires	H<F	H=F			
Employés et ouvriers qualifiés	H<F	H<F	H=F		
Professions peu qualifiées	H<F	H<F	H<F	H=F	
Indépendants	-	-	-	H>F	H=F

« H=F » indique que les groupes socioéconomiques des deux conjoints sont considérés comme approximativement de même statut. « H<F » indique que le statut social de la femme est supérieur à celui de son conjoint (hypogamie féminine), alors que H>F indique la situation inverse (hypergamie féminine). « - » indique l'absence de classement.

La partie supérieure de la table est l'inverse de sa partie inférieure.

2.3 Champ : les jeunes couples cohabitants

Tous les couples cohabitants, mariés ou non, sont retenus. Contrairement à la restriction adoptée dans le cas de la France par cohérence avec les anciennes séries d'enquête, plusieurs couples peuvent être identifiés dans un ménage. Cette précision des données s'avère plus utile ici que dans le cas français : la proportion de ménages comprenant plusieurs couples dépasse 10 % dans certaines régions de Bulgarie, de Pologne et de Roumanie, avec un maximum à 20 % dans ce dernier pays (voir colonne « Ménages multiples » dans le tableau 8.3, p. 496).

Nous limitons l'analyse aux jeunes couples, dans lesquels l'un des conjoints est âgé de 20 à 39 ans, dans l'objectif de mettre en relation au chapitre suivant l'homogamie avec des caractéristiques observables dans les années récentes.

L'échantillon final est constitué de 230 000 couples pour l'homogamie d'éducation, et de 176 000 couples pour l'homogamie socioéconomique. Sur 64 régions, les effectifs sont supérieurs à 1000 couples dans 55 régions pour l'éducation, et dans 46 pour le groupe

socioéconomique (Tableau 7.5, p. 454). Ils sont toujours supérieurs à 500, à l'exception d'une région pour l'éducation (Madrid) et de cinq pour le groupe socioéconomique (Bruxelles, Allemagne de l'Est, Madrid, Irlande du Nord et Pays de Galles, Écosse).

Pour finir, notons que, comme dans les analyses portant sur la France (chapitre 2), la modélisation se fonde sur les tables pondérées. La somme des effectifs de la table est égale au nombre réel de couples dans l'échantillon total, mais chaque région est représentée à proportion de sa population dans l'Union européenne. Le plan de sondage complexe des multiples dispositifs nationaux constituant l'Enquête européenne sur les forces de travail, non disponible dans les données, n'a pas pu être pris en compte pour le calcul des erreurs-types.

3 La stratégie de modélisation : le modèle à forme de régression multiple

3.1 Modéliser les variations de l'homogamie

La modélisation des variations de l'homogamie relative entre régions européennes présente des similarités avec celle des variations temporelles dans le cas de la France, mais aussi des différences importantes. Comme dans les chapitres précédents (Encadré 2.1, p. 136), les tables d'homogamie à analyser comportent trois dimensions, avec en lignes le groupe d'appartenance de l'homme, en colonnes celui de la femme, la troisième dimension (couche) correspondant à la région. Le modèle d'indépendance conditionnelle postule ainsi que le choix du conjoint est totalement aléatoire une fois tenu compte de la structure de la population de chaque région. Le modèle de stabilité de l'association revient quant à lui à affirmer que la structure et le niveau de l'homogamie sont les mêmes dans toutes les régions une fois contrôlées les différences de structure : ce modèle correspond donc à l'application à l'homogamie de l'hypothèse Featherman-Jones-Hauser présentée plus haut.

Au-delà de ces modèles de référence, la modélisation des variations entre régions s'éloigne quelque peu de celle au fil des années. En effet, comme on s'en rendra compte ci-dessous, les différences entre pays sont nettement plus marquées que celles au cours du temps au sein d'un même pays. Alors que les évolutions temporelles sont relativement lentes, et suivent généralement des tendances régulières sur des décennies, menant la

structure de l'association d'un état de départ à un état d'arrivée, les contrastes entre pays ne peuvent se résumer en une simple opposition entre deux pôles. En outre, le recours au modèle d'association lignes-colonnes RC-L pour décrire l'association parmi les configurations exogames laisserait ouverte la question de la description des configurations endogames, essentielle dans l'homogamie. De ce fait, la stratégie retenue au chapitre 2 n'est plus adaptée ici : une autre approche est nécessaire.

Pour ce faire, nous introduisons une extension du modèle log-multiplicatif à forme de régression (Goodman et Hout, 1998) déjà utilisé aux chapitres 2 et 4, où il était apparu comme la meilleure manière de résumer les évolutions de l'homogamie relative en France sur quarante ans. Ce modèle présente pour l'analyse des variations entre régions l'inconvénient de supposer que celles-ci se réduisent à une opposition entre deux structures-types de l'homogamie relative, l'ensemble des régions pouvant être décrites par leur degré de proximité à l'un ou l'autre modèle. La généralisation du modèle à forme de régression à plusieurs dimensions (Encadré 7.2) va nous permettre d'enrichir cette synthèse des différences régionales en supposant que les régions peuvent être décrites en référence à plus de deux structures-types. Cette généralisation, que nous désignerons sous le nom de modèle à forme de régression multiple, comprend plusieurs dimensions sur lesquelles les régions se voient attribuer une coordonnée indiquant dans quelle mesure leur structure de l'homogamie relative est proche de l'un ou l'autre pôle opposés sur la dimension considérée. À chaque pôle est associée une structure-type. Le croisement des multiples dimensions dessine ainsi un espace des régions européennes décrivant les oppositions entre ensembles de régions présentant des structures de l'homogamie similaires. À l'origine des axes correspond la situation moyenne des régions européennes.

À la différence des modèles d'association lignes-colonnes RC-L utilisés au chapitre 2, le modèle à forme de régression multiple décrit pour chaque dimension l'association complète entre caractéristiques des conjoints, à l'aide d'un paramètre (ou d'un *odds ratio* généralisé) par cellule. L'association correspondant à une région donnée est obtenue en prenant la somme de l'association moyenne (valable pour l'origine des axes) et de celles correspondant à chacune des dimensions du modèle. Plutôt que de retenir cette lecture abstraite des paramètres du modèle, nous représentons plus bas la structure de l'association correspondant à des régions typiques de chaque pôle, entre lesquelles les régions sont

placées sur les axes. Ces graphiques, sous la forme de chiasmogrammes (Bozon et Héran, 1987, 1988), consistent à représenter la table d'homogamie de chaque région-type en attribuant aux cellules une aire proportionnelle à leurs effectifs attendus en situation de mise en couple au hasard, et à représenter leurs *odds ratios* généralisés par des niveaux de gris. Les hauteurs des lignes et les largeurs des colonnes sont proportionnelles aux distributions marginales des groupes dans la population. Cette approche nous semble offrir une réponse satisfaisante à l'objectif d'étudier les variations de l'homogamie au-delà des contraintes structurelles sans pour autant perdre de vue l'influence de la structure de la population, qui n'intervient pas dans le modèle.

Lorsque notre attention porte sur le niveau plutôt que sur la structure de l'homogamie dans une région donnée, nous recourons comme précédemment à l'*odds ratio* absolu moyen (MAOR, cf. Annexe B, p. 525). Cet indicateur est calculé à partir des estimations du modèle. Nous verrons qu'il fournit une information presque parfaitement identique à celle du premier axe du modèle à forme de régression.

Encadré 7.2 : Équations du modèle log-multiplicatif à forme de régression multiple

Les premiers modèles utilisés ici ont été introduits au chapitre 2 (Encadré 2.1, p. 136). Nous ne présentons ici que l'extension à plusieurs dimensions du modèle à forme de régression.

En notant m_{hfr} les effectifs prédits par le modèle pour la cellule à l'intersection de la ligne h (catégorie de l'homme), de la colonne f (catégorie de la femme) et de la couche r (région), dans une table de dimensions $H \times F \times R$, ce modèle s'écrit :

$$\log m_{hfr} = \lambda + \lambda_h^H + \lambda_f^F + \lambda_r^R + \lambda_{hr}^{HR} + \lambda_{fr}^{FR} + \lambda_{hf}^{HF} + \sum_{m=1}^M \phi_{mr} \psi_{hfm}$$

M est le nombre de dimensions du modèle, ϕ_{mr} le coefficient de couche relatif à la dimension m pour la région r , ψ_{hfm} sont les coefficients d'interaction entre lignes et colonnes correspondant à la dimension m , et λ_{hf}^{HF} ceux ne dépendant d'aucune dimension. Lorsque $M = 0$, on retrouve le modèle de stabilité de l'association, et lorsque $M = 1$, le modèle à forme de régression dans sa forme originale (Goodman et Hout, 1998).

Comme dans ces deux modèles, les coefficients d'interaction entre lignes et colonnes sont rendus identifiables par une contrainte de sommes nulles en lignes et en colonnes :

$$\sum_{h=1}^H \lambda_{hf}^{HF} = \sum_{f=1}^F \lambda_{hf}^{HF} = \sum_{h=1}^H \psi_{hfm} = \sum_{f=1}^F \psi_{hfm} = 0 \quad \text{pour tout } m$$

Il est en outre nécessaire d'imposer deux contraintes supplémentaires à chaque dimension. Comme au chapitre 2 avec le modèle à forme de régression à une seule dimension, il est possible de fixer par convention le coefficient ϕ_{mr} à 0 pour l'une des couches, et à 1 pour une autre ; cependant, le choix de ces couches de référence n'est pas aussi évident dans le cas de régions que dans celui d'années d'enquête. Nous avons donc choisi ici de contraindre les coefficients de couche ϕ_{mr} à avoir une moyenne pondérée nulle (centrage), et de fixer l'écart-type des coefficients d'interaction ψ_{hfm} à 1 (réduction) :

$$\sum_{r=1}^R \phi_{mr} P_{++r} = 0 \quad \text{pour tout } m$$

$$\sum_{h=1}^H \sum_{f=1}^F \psi_{hfm}^2 = 1 \text{ pour tout } m$$

avec P_{++r} la proportion d'individus de l'échantillon dans la région r .

Ainsi, le coefficient ϕ_{mr} associé à une région donnée mesure l'écart-type des coefficients d'association entre lignes et colonnes sur cette dimension, ce qui est équivalent au coefficient d'association intrinsèque (cf. Annexe B, p. 525). Les coefficients d'interaction entre lignes et colonnes λ_{hf}^{HF} mesurent l'association moyenne de l'ensemble des couches.

Finalement, dans le cas où $M > 1$, deux contraintes interdimensionnelles supplémentaires sont requises pour chaque combinaison de dimensions. Nous imposons donc l'orthogonalité des coefficients de couche ϕ_{mr} ⁸ :

$$\sum_{r=1}^R \phi_{mr} \phi_{m'r} P_{++r} = 0 \text{ pour tout } m \neq m'$$

Comme dans les analyses factorielles et les modèles RC, les dimensions sont ordonnées de manière à ce que la première décrive au mieux la variance de l'association, et que les suivantes décrivent au mieux la variance résiduelle une fois tenu compte des dimensions précédentes.

Dans les résultats présentés dans le corps du texte, ces conventions n'affectent que les axes de la représentation graphique des coefficients de couche ϕ_{mr} . Les coefficients d'interaction entre lignes et colonnes correspondant à une région donnée ne dépendent pas de l'identification des dimensions.

Le modèle à forme de régression multiple à M dimensions présente, sans tenir compte des cellules parfois ignorées, $(H-1)(F-1)(R-1) - M[(H-1)(F-1) + (R-M-1)]$ degrés de liberté. Les *odds ratios* conditionnels relatifs à la région r s'expriment comme un produit (somme sur l'échelle logarithmique) de ceux correspondant à chaque dimension :

$$\theta_{hf, h'f'(r)} = \exp \left[\left(\lambda_{hf}^{HF} + \lambda_{h'f'}^{HF} - \lambda_{hf'}^{HF} - \lambda_{h'f}^{HF} \right) + \sum_{m=1}^M \phi_{mr} \left(\psi_{hfm} + \psi_{h'f'm} - \psi_{hf'm} - \psi_{h'fm} \right) \right]$$

8 En pratique, ces contraintes sont imposées à l'aide d'une décomposition en valeurs singulières de la matrice des coefficients d'interaction lignes-colonnes prédits par le modèle pour chaque région, après soustraction de la moyenne pondérée des coefficients pour l'ensemble des régions.

Pour finir, notons que le modèle à forme de régression multiple est une généralisation du modèle UNIDIFF, du modèle à forme de régression, mais aussi du modèle RC(M)-L. En effet, il est possible de contraindre la structure de l'interaction entre lignes et colonnes, représentée ici par les paramètres ψ_{hfm} , à être égale à un produit de scores en lignes et en colonnes (Goodman et Hout, 1998, p. 180). Ces scores étant centrés et réduits, ils satisfont aux contraintes déjà présentées, et le coefficient ϕ_{mr} devient le coefficient d'association intrinsèque, justifiant l'équivalence annoncée plus haut. Si les coefficients communs à toutes les couches λ_{hf}^{HF} sont par ailleurs contraints à être nuls, on se ramène au modèle RC(M)-L.

3.2 Modéliser les variations de l'hypergamie

L'analyse de l'hypergamie suit l'approche introduite au chapitre 4 (p. 258), en recourant à la matrice d'hypergamie socioéconomique présentée plus haut. Afin de nous concentrer sur les asymétries hommes-femmes, nous prenons comme modèle de base le modèle à forme de régression multiple retenu pour la modélisation de l'homogamie, mais en contraignant les paramètres d'association à être symétriques, et en ignorant les cellules de la diagonale de la table. L'introduction d'un paramètre⁹ mesurant la sur- ou sous-représentation des configurations dans lesquelles le statut de l'homme est supérieur à celui de sa conjointe nous permet d'identifier l'intensité moyenne de l'hypergamie ou de l'hypogamie dans l'ensemble de l'Europe au-delà de l'influence de la structure de la population de chaque région. Puis l'introduction d'un paramètre d'hypergamie spécifique à chaque région nous permet d'étudier les variations entre régions de cette intensité.

Cette stratégie doit être légèrement adaptée dans le cas de l'hypergamie d'éducation. Du fait du nombre très faible de catégories retenues (trois), une fois la diagonale de la table exclue, le modèle d'indépendance (contrôlant uniquement les marges de la table d'homogamie de chaque région) inclut déjà assez de paramètres pour décrire parfaitement l'association quasi-symétrique de chaque région. En d'autres termes, il n'est pas possible d'étudier la structure de l'association quasi-symétrique en dehors de la diagonale : chaque ligne et chaque colonne ne contenant que deux cellules prises en compte dans le modèle, les

⁹ Dans le cas du groupe socioéconomique, un paramètre supplémentaire mesure la sur- ou sous-représentation des configurations non ordonnées ; nous ne le reprenons pas dans la présentation des résultats.

paramètres d'interaction quasi-symétrique sont redondants. L'introduction d'un paramètre d'hypergamie est néanmoins possible. Lorsqu'un paramètre d'hypergamie spécifique à chaque région est introduit, on se ramène au modèle saturé, qui décrit parfaitement les données. Cette particularité, qui tient uniquement au peu de détail de la classification utilisée, ne pose aucun problème d'identification, d'interprétation ni de stabilité des paramètres.

Comme au chapitre 4, le paramètre d'hypergamie est exprimé sous la forme plus interprétable de l'*odds ratio* généralisé anti-symétrique (cf. Encadré 4.1, p. 259). Rappelons que cet indicateur est égal au rapport entre les *odds ratios* généralisés correspondant aux situations d'hypergamie féminine et ceux correspondant aux situations d'hypogamie féminine.

4 Les variations de l'homogamie

4.1 L'homogamie absolue : une correspondance entre composantes éducative et socioéconomique

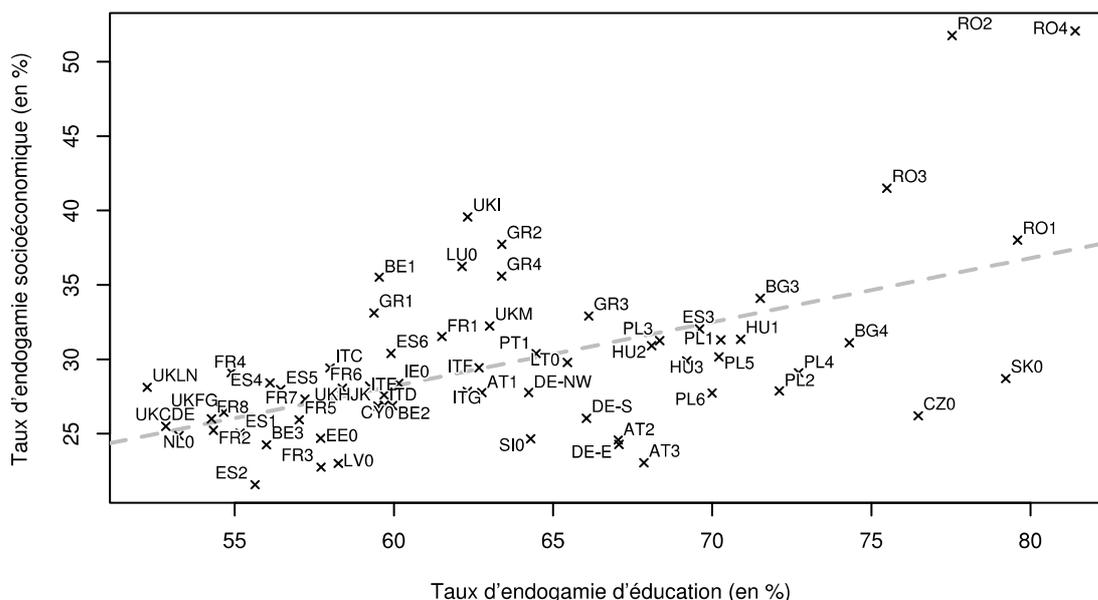
L'étude des taux absolus d'endogamie fait immédiatement apparaître à la fois des différences majeures entre régions d'Europe, et une correspondance assez forte entre homogamie d'éducation et homogamie socioéconomique (Figure 7.1). La corrélation entre ces deux indicateurs, qui sont extrêmement sensibles à la structure de la population de chaque région, est de 0,55. Ainsi, les deux dimensions de l'homogamie étudiées ici semblent bien relever au moins en partie du même phénomène – ce que nous tenterons de vérifier plus en détail dans le reste de ce chapitre et dans le suivant.

Les variations de l'endogamie d'éducation sont d'interprétation assez directe. Les couples associent le moins fréquemment deux personnes du même niveau d'éducation en Europe de l'Ouest (Royaume-Uni, Pays-Bas, France, Espagne, Belgique, Irlande, Italie du Nord et du Centre) et du Nord (Estonie, Lettonie) avec des taux inférieurs à 60 %¹⁰. Au contraire, l'endogamie est particulièrement élevée dans certaines régions d'Europe centrale et orientale, avec des taux supérieurs à 75 % en Roumanie, en Slovaquie et en République

10 Rappelons que le taux d'endogamie est un indicateur extrêmement sensible au degré d'agrégation utilisé (cf. chapitre 1), ce qui explique que l'endogamie d'éducation apparaisse ici très élevée même dans les régions les moins endogames.

tchèque, et supérieurs à 70 % en Bulgarie, en Pologne et dans une région de Hongrie (Budapest, HU1).

Figure 7.1 : Taux d'endogamie d'éducation et socioéconomique dans les régions d'Europe



Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour le groupe socioéconomique)

La ligne pointillée correspond à la droite de régression.

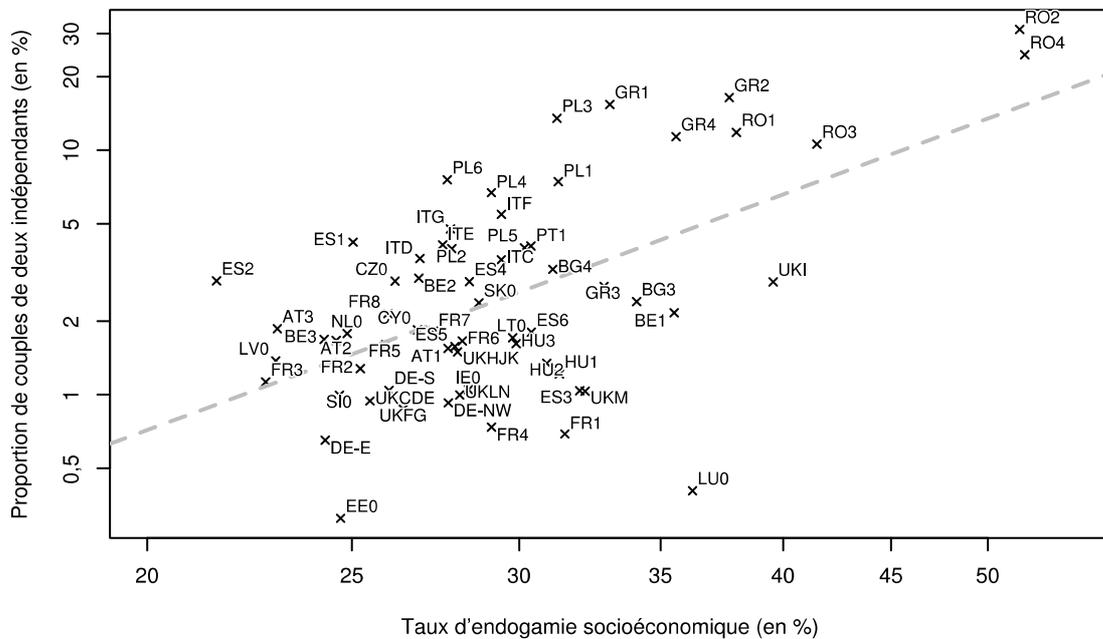
Cf. Tableau 7.5 (p. 454) pour la correspondance entre codes et noms des régions.

Globalement, il apparaît donc clairement que l'endogamie d'éducation diminue avec le niveau de développement¹¹. En revanche, comme exception à cette règle, elle est nettement plus forte dans les capitales lorsqu'elles sont séparées dans une région à part que dans les régions qui les entourent : c'est le cas à Budapest (HU1), Bruxelles (BE1), Londres (UKI), Madrid (ES3) ou encore Paris (FR1). Cette différence s'explique en partie par la plus forte proportion de couples formés de deux diplômés du supérieur dans ces régions au niveau d'éducation plus élevé que la moyenne.

11 On trouvera le produit intérieur brut par habitant des différentes régions dans la colonne « PIB/hab. » du tableau 8.3 (p. 496).

Les variations de l'endogamie socioéconomique sont encore plus marquées. Les régions les moins endogames (taux inférieur à 30 %) sont celles déjà citées d'Europe de l'Ouest et du Nord, auxquelles se rattachent ici l'Autriche, l'Allemagne, la Slovénie, la République tchèque, la Slovaquie, la Pologne et la Lituanie. L'endogamie la plus forte s'observe encore une fois dans les deux régions les plus rurales de Roumanie (à 50 %), ainsi que dans le reste de ce pays, en Grèce, mais aussi dans plusieurs capitales occidentales – Luxembourg (LU0), Bruxelles (BE1) et Londres (UKI) – qui présentent des taux supérieurs à 35 %.

Figure 7.2 : Taux d'endogamie socioéconomique et proportion de couples de deux indépendants dans les régions d'Europe



Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour le groupe socioéconomique)

La ligne pointillée correspond à la droite de régression. Une échelle logarithmique est utilisée afin d'augmenter la lisibilité du graphique.

Cf. Tableau 7.5 (p. 454) pour la correspondance entre codes et noms des régions.

Ces deux ensembles de régions – zones très rurales et capitales – sont situés aux deux extrêmes de l'échelle de développement économique en Europe et présentent des taux

d'endogamie d'éducation très éloignés (à l'exception de la Grèce). Leurs forts taux d'endogamie socioéconomique ont des sources très différentes. Pour les premières, ils proviennent presque exclusivement de la forte proportion de couples formés de deux indépendants, et notamment de deux agriculteurs (Figure 7.2). Pour les secondes, ils s'expliquent par celle des couples dont les deux membres sont cadres ou professions libérales et intellectuelles. Nous verrons plus bas que la particularité de ces deux types de régions persiste en partie une fois contrôlée la structure de la population de chaque région.

4.2 L'homogamie relative : portrait de huit régions-types

L'application de la stratégie de modélisation décrite à la section précédente aux tables d'homogamie d'éducation et de groupe socioéconomique conduit à des conclusions très proches concernant la sélection du meilleur modèle (Tableau 7.3). Le modèle d'indépendance conditionnelle, qui suppose l'absence d'une tendance à l'homogamie relative, classe correctement plus des trois quarts des couples, signe comme dans le cas de la France de l'importance des contraintes structurelles dans la détermination du choix du conjoint. Le modèle de stabilité de l'association, qui postule que l'homogamie relative est la même dans toutes les régions européennes, améliore de manière très nette la description des données, en faisant monter cette proportion au-dessus de 90 %. Ce résultat, valable aussi bien pour l'éducation que pour le groupe socioéconomique, indique que l'hypothèse Featherman-Jones-Hauser possède une certaine pertinence appliquée à l'homogamie : les proximités et distances entre groupes sociaux du point de vue du choix du conjoint présentent de grandes similarités entre régions européennes.

Néanmoins, nos données nous permettent d'aller au-delà de ce constat en mettant en évidence des différences entre régions. Le modèle UNIDIFF, qui suppose que la structure de l'association est commune, mais que les régions diffèrent par le degré d'intensité de l'homogamie relative, améliore très nettement la description des données selon les critères BIC et AIC. Le gain descriptif est cependant modeste : 2,8 % des couples dans le cas de l'homogamie d'éducation, et seulement 0,6 % dans le cas de l'homogamie socioéconomique.

Tableau 7.3 : Indicateurs d'ajustement des modèles concernant l'homogamie

	D. L.	Déviante	Δ (%)	BIC	AIC
Éducation :					
Indépendance conditionnelle	256	72 778	22,51	69 618	72 266
Stabilité	252	6 814	5,44	3 703	6 310
UNIDIFF	189	1 805	2,66	-529	1 427
Forme de régression, 1 dimension	186	1 442	2,36	-854	1 070
Forme de régression, 2 dimensions	122	613	1,34	-893	369
Forme de régression, 3 dimensions	60	189	0,48	-551	69
Groupe socioéconomique :					
Indépendance conditionnelle	2 304	53 157	21,03	25 325	48 549
Stabilité	2 268	10 096	8,04	-17 302	5 560
UNIDIFF	2 205	8 824	7,41	-17 813	4 414
Forme de régression, 1 dimension	2 170	8 217	7,20	-17 996	3 877
Forme de régression, 2 dimensions	2 074	6 956	6,40	-18 098	2 808
Forme de régression, 3 dimensions	1 980	6 022	5,94	-17 896	2 062

D. L. : degrés de liberté. Δ : indice de dissimilarité.

Enfin, relâchant l'hypothèse d'une structure commune de l'homogamie relative en Europe, le modèle à forme de régression suppose que l'association observée dans chaque région peut se résumer par référence à plusieurs régions-types. Le modèle à une seule dimension ne retient que deux régions de référence ; quoiqu'il soit préféré au précédent par le BIC et l'AIC, il ne permet une réduction de l'indice de dissimilarité que de respectivement 0,3 et 0,2 points de pourcentage. L'introduction d'une deuxième dimension est elle aussi justifiée par une réduction du BIC et de l'AIC ; le gain descriptif est plus net, avec respectivement 1,0 et 0,8 points de pourcentage¹². Finalement, l'introduction d'une troisième dimension permet d'améliorer encore la description des données de respectivement 0,9 et 0,4 points de pourcentage. Quoique l'AIC soit en faveur de l'ajout de cette troisième dimension, le BIC préfère le modèle à seulement deux dimensions. Pour limiter la complexité de l'interprétation des résultats, nous nous en tenons à deux dimensions.

12 Notons que l'introduction d'une seconde dimension consomme nettement plus de degrés de liberté, par rapport au modèle à forme de régression à une dimension, que le passage du modèle UNIDIFF à ce dernier modèle. L'ajout d'une dimension supplémentaire est en fait (aux contraintes interdimensionnelles près) équivalent au passage du modèle de stabilité au modèle UNIDIFF.

Cette décision n'a qu'un effet négligeable sur la mesure des variations dans le niveau de l'homogamie en Europe : l'intensité de l'homogamie relative d'éducation et de groupe socioéconomique décrite par le modèle à deux dimensions, telle que mesurée par l'*odds ratio* absolu moyen, présente une corrélation de 0,9 (sur l'échelle logarithmique) avec celle décrite par le modèle à trois dimensions. Nous commençons par présenter les variations géographiques de cette intensité, avant d'analyser plus en détail celles de la structure de l'homogamie.

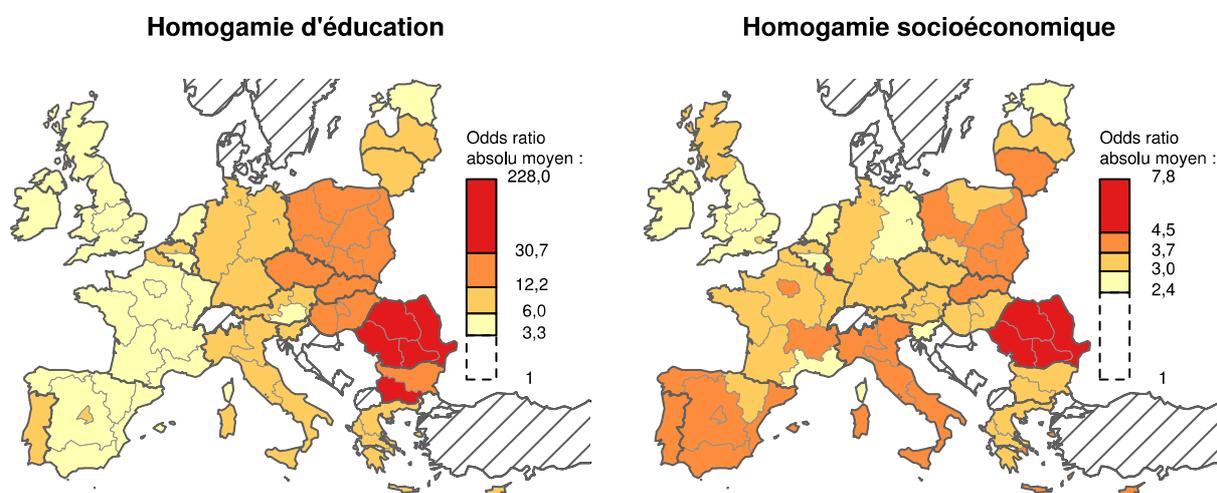
4.2.1 La géographie de l'homogamie relative en Europe

La représentation sur une carte des *odds ratios* absolus moyens mesurant l'intensité de l'homogamie d'éducation et de groupe socioéconomique estimée par le modèle à forme de régression à deux dimensions fait apparaître des écarts très importants entre régions (Figure 7.3 ; les valeurs exactes sont reprises dans le tableau 8.3, p. 496). Deux résultats fondamentaux apparaissent immédiatement.

Le premier tient à la corrélation relativement forte entre homogamie d'éducation et homogamie socioéconomique (coefficient de 0,59), associée à une supériorité systématique de l'homogamie d'éducation. Dans 61 régions sur 64, c'est l'homogamie d'éducation qui est la plus forte, en dépit du fait que le groupe socioéconomique est mesuré de manière plus précise¹³. L'*odds ratio* absolu moyen dépasse largement 30 dans de nombreuses régions du point de vue de la première composante, alors que les valeurs les plus élevées demeurent inférieures à 8 pour la seconde. L'importance comparable des aspects éducatifs et socioéconomiques dans le choix du conjoint, relevés dans le cas de la France au chapitre 2, n'est ainsi pas une constante au niveau européen. Elle est spécifique aux pays de l'Ouest du continent, et semble être conditionnée à une relative faiblesse de l'homogamie d'éducation.

13 Rappelons que contrairement à sa version absolue, la mesure relative de l'intensité de l'homogamie ne croît pas mécaniquement lorsqu'une classification plus agrégée est utilisée. On obtient des résultats similaires avec une classification plus détaillée du niveau d'éducation.

Figure 7.3 : Distribution géographique de l'homogamie relative d'éducation et de groupe socioéconomique en Europe



Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour l'homogamie socioéconomique)

Une échelle logarithmique est utilisée, les écarts sont donc équivalents à des rapports entre *odds ratios*. Le découpage en classes a été réalisé automatiquement à l'aide de l'algorithme de Jenks (1967), mais la classe comprenant uniquement deux régions de Roumanie a été fusionnée avec celle immédiatement en dessous du fait de l'incertitude portant sur la réalité de cette différence (note 14, p. 430).

Le second résultat tient à la forte corrélation entre mesures absolue et relative de l'homogamie. La corrélation entre le taux d'endogamie absolue déjà décrit à la figure 7.1 (p. 424) et l'*odds ratio* absolu moyen présenté dans la figure 7.3 est de 0,88 concernant l'éducation et 0,69 concernant le groupe socioéconomique. Nous retrouvons ici à propos de l'homogamie le résultat établi par A. Esteve, J. García-Román et I. Permanyer (2011), évoqué au chapitre 4, selon lequel hypergamie absolue et relative varient généralement dans le même sens lorsqu'un grand nombre de pays sont comparés. L'influence de la structure de la population et des couples semble donc s'étendre au-delà des contraintes de disponibilité des conjoints, façonnant plus largement les mécanismes de choix du conjoint.

Néanmoins, cette relation doit être relativisée : la Roumanie apparaît comme un pays à part dans l'Union européenne du point de vue de l'intensité de l'homogamie selon les deux

composantes étudiées, ce qui renforce artificiellement la corrélation entre les deux mesures¹⁴. Lorsque les régions de Roumanie sont exclues de l'échantillon, la corrélation entre homogamie absolue et homogamie relative tombe à 0,39 pour le groupe socioéconomique, mais se maintient au même niveau pour l'éducation. Le lien entre homogamie socioéconomique absolue et relative semble donc tenir largement aux variations de la proportion d'agriculteurs, nettement plus élevée en Roumanie que dans le reste de l'Europe.

La distribution géographique de l'homogamie relative d'éducation est assez simple, avec une décroissance d'Est en Ouest, à quelques exceptions près. Elle est ainsi très élevée dans les anciennes démocraties populaires : Roumanie, Bulgarie, Slovaquie, République tchèque, Hongrie, Pologne et Slovénie. Elle est moins forte dans une bande située au Sud, à l'Ouest et au Nord de cet ensemble, soit respectivement à Chypre, en Grèce, en Italie, en Autriche, en Allemagne, mais aussi en Lituanie, en Lettonie et en Estonie (pays anciennement membres de l'Union soviétique). Enfin, elle est la plus faible en Europe occidentale, à l'exception du Portugal : Pays-Bas, France, Royaume-Uni, Irlande, Espagne – la Belgique étant entre les deux dernières zones. Il est intéressant de noter que le Portugal fait aussi exception en Europe par le très faible niveau d'éducation de sa population.

La répartition de l'homogamie socioéconomique est moins régulière. Outre la Roumanie, qui se distingue encore nettement des autres pays par sa très forte homogamie, on peut relever une certaine tendance à une plus forte homogamie dans l'Est et dans le Sud de l'Europe (mais pas dans l'Europe du Sud-Est, à l'exception notable de la Roumanie) : Lituanie, Pologne, Slovaquie, Italie, Espagne, Portugal, ainsi que la Belgique et certaines régions de France. Plus qu'une régularité géographique approximative, les pays cités ont pour point commun d'être ceux dans lesquels la religion catholique est la plus fortement implantée – à l'exception de l'Irlande, à dominante catholique mais dont l'homogamie

14 Soulignons cependant que le modèle à forme de régression, s'il décrit assez précisément la table d'homogamie observée dans l'Est et l'Ouest de la Roumanie (RO2 et RO4, avec des indices de dissimilarité de respectivement 0,8 % et 2,0 %), semble surestimer le niveau des *odds ratios* absolus moyens de ces régions : ceux-ci dépassent 200, alors que le calcul à partir des données brutes fournit des valeurs plus proches des régions voisines, à respectivement 84 et 54. Cette divergence s'explique par le fait que les couples associant les deux extrêmes de l'échelle sociale sont extrêmement rares dans ces régions (les deux cellules correspondantes de chaque table réunissent chacune moins de quatre individus observés) : un écart d'un individu suffit à faire varier l'*odds ratio* de manière très importante. Ce phénomène illustre la difficulté à mesurer précisément des *odds ratios* extrêmes, qui impliquent nécessairement un fort déséquilibre entre cellules et donc des effectifs extrêmement faibles dans certaines d'entre elles.

socioéconomique est parmi les plus faibles. Nous reviendrons plus en détail au chapitre suivant sur les déterminants religieux et culturels de l'homogamie, ainsi que sur le lien potentiel entre activité féminine (dont la religion serait l'un des facteurs) et homogamie socioéconomique.

Ce premier aperçu des variations de l'intensité de l'homogamie en Europe montre d'ores et déjà que, si les aspects éducatif et socioéconomique sont en grande partie liés, leur corrélation n'est pas systématique. L'intensité de l'homogamie relative d'éducation semble à première vue très négativement corrélée au niveau de développement d'une région et au caractère agricole de sa population. Cette relation apparaît moins nette dans le cas de l'homogamie socioéconomique, comme l'illustre de manière frappante l'*odds ratio* absolu moyen très élevé du Luxembourg (LU0).

D'une manière générale, nous retrouvons une homogamie relative en termes socioéconomique plus élevée dans les grandes capitales lorsqu'elles sont séparées dans une région propre, que dans les régions qui les entourent : c'est le cas de Londres, Bruxelles (non visible sur la carte), Paris et l'Île-de-France (la région entourant Lyon pourrait être rattachée à ce schéma). Le fait que cette particularité des capitales ne se retrouve pas dans le cas de l'homogamie d'éducation pourrait tenir au manque de détail de la classification en trois niveaux utilisée dans ce dernier cas. Cette observation va à l'encontre de la vision de la ville comme le lieu de mélanges entre les groupes sociaux qui était celle des auteurs classiques de la sociologie du début du XX^e siècle. Nous pouvons faire l'hypothèse que dans des grandes villes, chaque groupe socioéconomique est assez grand pour que les individus n'aient pas besoin d'aller chercher un conjoint en dehors de leur groupe d'appartenance : en d'autres termes, des marchés conjugaux locaux plus développés facilitent la satisfaction des préférences individuelles à l'intérieur de son propre groupe. Il est aussi possible que les inégalités économiques plus fortes qui caractérisent les capitales soient un frein aux contacts entre groupes. Nous reviendrons sur ces questions au chapitre suivant.

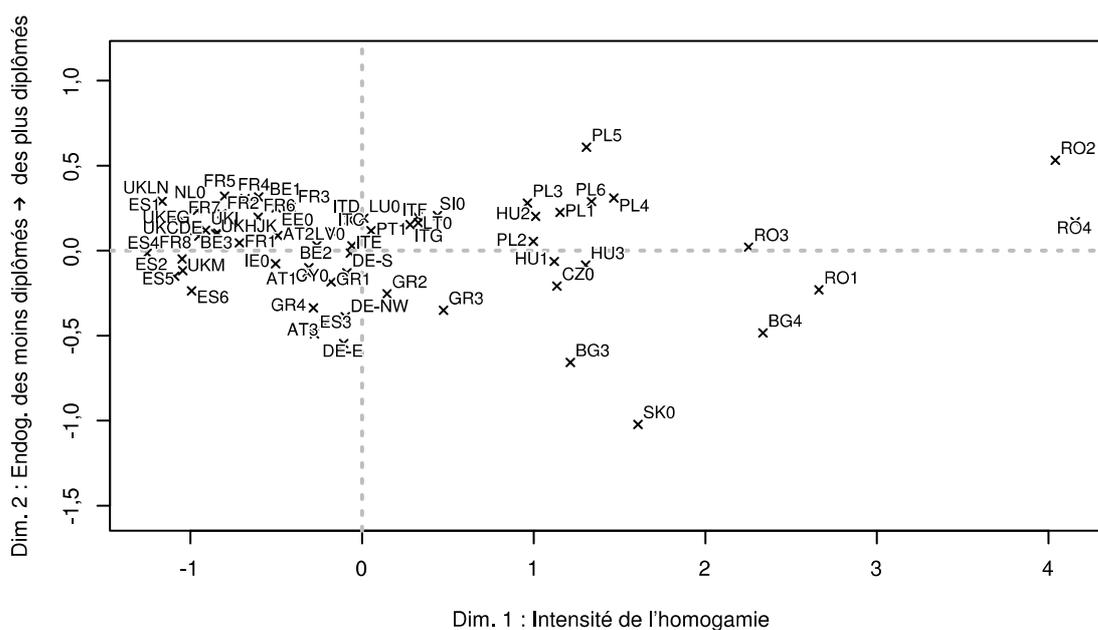
4.2.2 La structure de l'homogamie d'éducation : l'influence de l'organisation du système scolaire

La situation des régions européennes sur les deux dimensions du modèle à forme de régression retenu pour l'homogamie d'éducation est présentée dans la figure 7.4. Le premier

axe (horizontal) correspond à l'intensité globale de l'homogamie (corrélation presque parfaite avec l'*odds ratio* absolu moyen présenté à la figure 7.3).

Le second axe (vertical) introduit une distinction nouvelle, de moindre importance, qui oppose notamment deux groupes de pays anciennement communistes : les régions les plus rurales de Pologne (notamment le Sud-Ouest, PL6) et de Roumanie (Est, RO2, et Sud-Ouest, RO4) d'un côté, la Slovaquie et la Bulgarie de l'autre. Il distingue les régions dans lesquelles l'endogamie est plus forte en haut de l'échelle des diplômés, et dans une moindre mesure en son milieu (en haut de l'axe), de celles dans lesquelles elle est plus marquée en bas de cette échelle (en bas de l'axe).

Figure 7.4 : Dimensions du modèle à forme de régression multiple pour l'homogamie d'éducation



Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans
Cf. Tableau 7.5 (p. 454) pour la correspondance entre codes et noms des régions.

Nous illustrons le sens de ces deux oppositions en présentant la structure de l'homogamie relative dans quatre¹⁵ régions-types (Figure 7.5). Chaque chiasmogramme (voir

15 À s'en tenir à la spécification du modèle, seules trois structures de l'association doivent être décrites : une

p. 419) reprend les proportions marginales des hommes et des femmes par niveau d'éducation de la région prise pour exemple. Rappelons néanmoins que cette information, qui permet de replacer l'homogamie relative dans son contexte structurel, n'intervient pas dans la construction des deux axes du modèle à forme de régression, et ne représente donc pas nécessairement les régions situées à proximité de l'exemple choisi : seule la proximité en termes d'*odds ratios* est prise en compte.

L'arrangement des quatre chiasmogrammes de la figure 7.5 reprend sur chaque ligne l'opposition correspondant à un axe de la figure 7.4. Ainsi, la première ligne recoupe l'axe horizontal, et oppose le centre de l'Espagne (ES4, ne comprenant pas la région de Madrid) à l'Est de la Roumanie (RO2). La seconde ligne correspond à l'axe vertical, et oppose le Sud-Ouest de la Pologne (PL6) à la Slovaquie (SK0).

La première opposition correspond à une différence en termes d'intensité de l'homogamie. Celle-ci ne concerne pas le groupe des diplômés du secondaire complet, dont l'endogamie et les différentes configurations exogames sont assez proches dans les deux régions. Ce sont donc exclusivement les quatre *odds ratios* généralisés situés dans les angles de la table d'homogamie, qui mesurent la sur- ou sous-représentation des configurations associant un diplômé du supérieur à un peu diplômé, ou deux membres de ces groupes (endogamie), qui varient de manière radicale entre les deux régions-types. Ainsi, l'*odds ratio* généralisé mesurant l'endogamie relative des diplômés du supérieur est dans l'Est de la Roumanie de 1207, et celui mesurant l'endogamie relative des peu diplômés, de 2314. Dans le Centre de l'Espagne, ces valeurs sont respectivement de 4 et de 6. La structure de l'homogamie en Roumanie (et notamment dans ses régions les plus rurales), mais aussi moins nettement en Bulgarie, en Slovaquie, en Hongrie, en République tchèque et en Pologne, se caractérise ainsi par une endogamie élevée aux deux extrêmes de la distribution de l'éducation, en décalage avec la très grande ouverture du groupe intermédiaire majoritaire. Au contraire, en Espagne, dans la plupart des régions d'Europe occidentale, mais aussi en Allemagne de l'Est, en Estonie et en Lettonie, ces deux groupes extrêmes présentent un comportement bien moins marqué, se rapprochant de celui du groupe

structure moyenne, et une structure pour chacune des deux dimensions du modèle. Cependant, une telle présentation n'offrant aucune illustration concrète du résultat de la combinaison de ces composantes, il nous a semblé plus approprié de décrire les quatre positions extrêmes.

intermédiaire. On voit que cette opposition ne recoupe pas complètement la césure entre anciens blocs de l'Est et de l'Ouest, même si elle la suit en grande partie.

La seconde dimension, qui oppose le Sud-Ouest de la Pologne (PL5) à la Slovaquie (SK0), est d'importance moindre. De fait, ces régions-types sont toutes deux situées nettement à droite du premier axe, et partagent donc en partie la très forte endogamie des groupes extrêmes observée dans l'Est de la Roumanie (RO2). La différence entre elles tient au rapport entre endogamie des diplômés du supérieur et endogamie des peu diplômés. Ainsi, en Slovaquie, les *odds ratios* généralisés mesurant l'endogamie de ces deux groupes sont respectivement de 71 et 265, soit un rapport de 1 à 4¹⁶. À l'inverse, dans le Sud-Ouest de la Pologne, ces *odds ratios* sont de 127 et 48, soit un rapport de 3 à 1. En outre, l'endogamie des diplômés du secondaire complet est un peu plus élevée dans la première région (*odds ratio* généralisé de 4) que dans la seconde (1,1 seulement).

Cette opposition semble devoir être rapportée à la structure des inégalités au sein de chaque système scolaire, ou en d'autres mots, à la position des paliers de sélection principaux. En bas de l'axe, on trouve en effet des pays dont le système scolaire se divise précocement en filières hiérarchisées (Kogan, Gebel et Noelke, 2008, p. 14 ; Müller et Kogan, 2010, p. 226). En Slovaquie, le second cycle du secondaire est fortement stratifié entre une filière générale (*Gymnasium*), deux filières professionnelles et une filière courte d'apprentissage, les taux d'abandon étant nettement plus élevés dans les deux derniers groupes (Košta et Bednárík, 2008, p. 326). En République tchèque, les élèves peuvent passer un examen pour entrer au *Gymnasium* à partir de 11 ans ; dans tous les cas, l'accès au second cycle du secondaire est conditionné au passage d'un examen d'entrée spécifique à chaque filière (Straková, 2008, p. 219). Enfin, en Allemagne, les élèves sont orientés à la sortie de l'école primaire selon leurs résultats, et ceux qui suivent la moins prestigieuse des trois filières (*Hauptschule*) ne poursuivent pas toujours leur scolarité (par le biais de l'apprentissage) jusqu'à la fin du second cycle du secondaire (Schneider, 2008c, p. 98). Du fait de cette orientation précoce, les inégalités selon l'origine sociale dans l'orientation vers des filières qui ne mènent pas à des études supérieures sont plus élevées en République tchèque, en Allemagne de l'Est et en Hongrie que dans d'autres anciens pays communistes

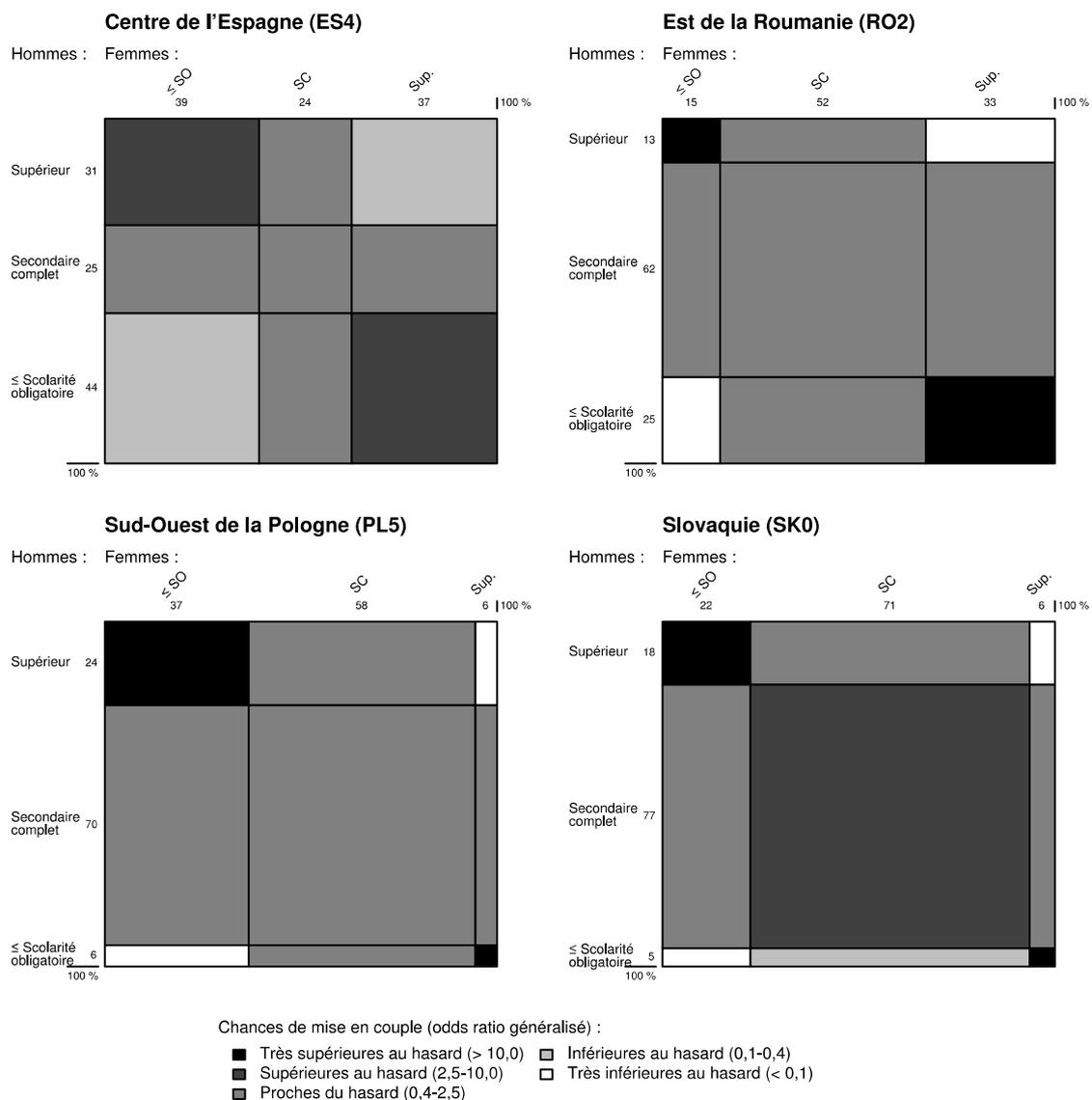
16 Cette forte spécificité de la Slovaquie avait déjà été relevée par H. Domański et D. Przybysz (2007, p. 518) à partir de données de l'Enquête sociale européenne. Une comparaison plus systématique avec cette étude n'est cependant pas possible du fait des différences concernant la classification utilisée.

comme la Pologne (Kogan, Gebel et Noelke, 2012 ; la Slovaquie et la Bulgarie ne faisaient pas partie des pays étudiés).

Outre ces pays dans lesquels le système éducatif est stratifié très tôt, se trouvent du même côté du second axe des pays dans lesquels la différenciation des filières est moins précoce, mais où l'interruption des études avant la fin du secondaire est relativement fréquente. C'est le cas de la Grèce, pays dans lequel les examens d'entrée dans le second cycle du secondaire ont été abolis en 1976 (Kappi et Stratoudaki, 2004, p. 306), mais où persiste un taux important de personnes n'ayant pas validé l'enseignement secondaire complet (autour de 30 % des individus âgés de 20 à 39 ans en 2011). L'Espagne correspond aussi, à un moindre degré, à cette description, une partie importante des individus n'ayant pas suivi le premier cycle d'études secondaires (Ortiz, 2008, p. 139).

Enfin, la position de la Bulgarie et de certaines régions de Roumanie semble devoir être attribuée aux inégalités scolaires dont souffre la forte minorité Rom dans ces deux pays (European Union Agency for Fundamental Rights, 2012, p. 15). En Bulgarie, pays à la proportion de Roms la plus élevée de l'Union européenne, le niveau d'éducation des minorités, et en particulier celui des Roms, est nettement inférieur à la moyenne, et les décrochages scolaires plus fréquents (Kostova, 2008, p. 171, 174). En Roumanie, la région de l'Est et du Centre (RO1) se trouve en bas du second axe, contrairement au reste du pays ; or, cette région est celle dans laquelle les Roms sont les plus nombreux. La forte endogamie des moins diplômés au sein de ces régions serait alors le reflet d'une endogamie ethnique autant que du fonctionnement du système scolaire.

Figure 7.5 : Structure de l'homogamie d'éducation dans quatre régions-types



Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans

Dans les quatre chiasmogrammes, les hauteurs des lignes et les largeurs des colonnes correspondent aux proportions d'hommes et de femmes dans chaque catégorie parmi les personnes en couple. L'aire des cellules est égale à la proportion attendue en situation de mise en couple aléatoire, alors que la teinte de gris mesure l'écart à cette référence (*odds ratio* généralisé).

Au contraire des trois ensembles de pays décrits ci-dessus, les régions du haut du second axe se caractérisent par une absence de paliers d'orientation jusqu'à un niveau relativement élevé conjuguée à un taux de décrochage assez faible. C'est notamment le cas du système polonais (Mach et Kryszczuk, 2008 ; Le Donné, 2014). Avant la réforme de 1999 (qui ne s'est appliquée qu'aux individus âgés de moins de 25 ans en 2011), une école primaire unique menait jusqu'à la fin du premier cycle secondaire à 15 ans, âge auquel lui succédaient trois types d'établissements du second cycle secondaire, dont seuls le lycée général et – après 1993 – le lycée technique, préparaient à l'examen d'entrée à l'université (*Matura*). Les trois voies menaient cependant toutes à la fin des études secondaires à 18, 19 ou 20 ans. Depuis la réforme, le tronc commun s'étend jusqu'à l'âge de 16 ans, les élèves devant ensuite choisir entre quatre filières qui fonctionnent sur un principe similaire au système précédent. La sélection ne s'opère donc pas de manière aussi précoce que dans les pays cités au paragraphe précédent, mais elle apparaît à l'âge de 15 ou 16 ans et détermine fortement les chances de poursuite d'études dans le supérieur. Cette caractéristique peut expliquer le décalage entre une faible endogamie des moins diplômés et une forte endogamie des diplômés du supérieur.

Le système éducatif roumain présente des caractéristiques similaires (Mocanu, 2008), ce qui semble expliquer la position en haut de l'axe des régions dans lesquelles les Roms sont peu représentés. En effet, la scolarité y est obligatoire jusqu'à 14 ans (15 ans depuis 1999) et le premier cycle du secondaire est unique. Une division en trois filières apparaît au second cycle du secondaire seulement, et seule la filière générale permet l'entrée dans le supérieur (nettement sous-développé jusqu'aux années 2000) après passage d'un examen. Les élèves issus des autres filières doivent compléter leur formation en intégrant la voie générale pour pouvoir rejoindre le supérieur, que ce soit dans une formation professionnelle ou générale. Ici aussi, la sélection la plus forte intervient donc à la fin du secondaire.

Finalement, même si ces pays se détachent moins sur le second axe, on peut citer les cas de la Slovénie et de l'Estonie, ainsi que du Royaume-Uni, dont les systèmes éducatifs sont eux aussi stratifiés de manière tardive (Müller et Kogan, 2010, p. 226).

Le manque de détail de notre classification des niveaux d'éducation nous interdit de pousser plus loin l'étude des différences en termes de structure de l'homogamie relative en Europe. Une analyse plus fine, potentiellement restreinte à un ensemble de pays pour

lesquels des données plus précises sont disponibles, permettrait de mettre à l'épreuve les interprétations que nous venons de proposer, en identifiant les seuils les plus significatifs de chaque système scolaire. On pourrait par exemple s'attendre à observer une forte endogamie des différentes filières du secondaire dans les régions dont le système scolaire est précocement différencié, particularité qui est invisible ici du fait de l'absence de distinction entre ces catégories.

4.2.3 La structure de l'homogamie socioéconomique

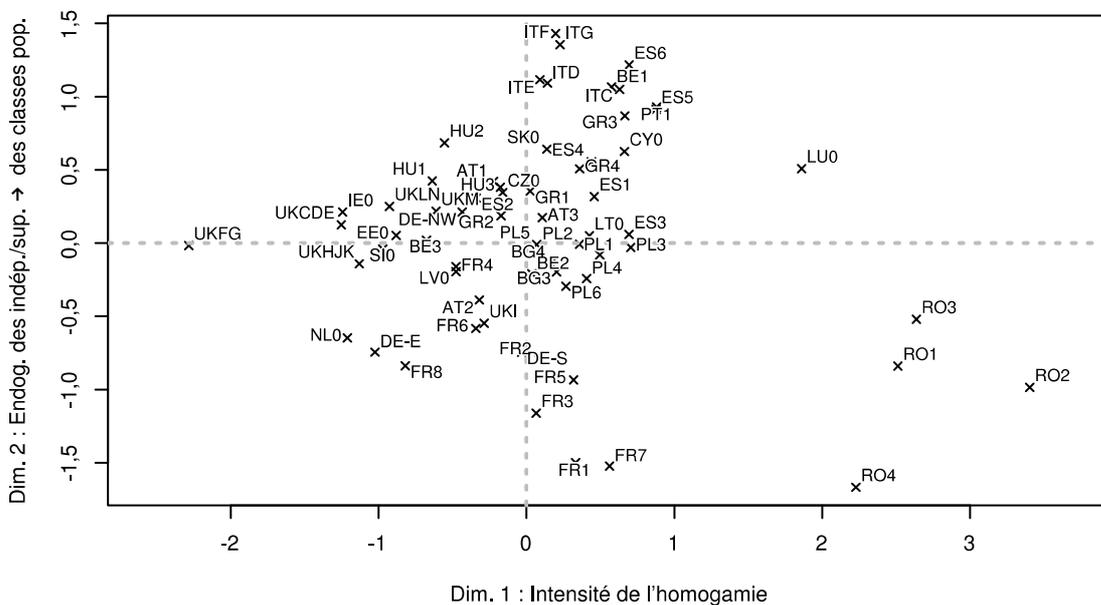
Les variations de l'homogamie socioéconomique se répartissent de manière plus équilibrée entre les deux dimensions du modèle à forme de régression (Figure 7.6). La première oppose les régions roumaines à celles du Nord-Est de l'Europe, dont en particulier celles du Royaume-Uni, ainsi que l'Irlande et les Pays-Bas ; comme pour l'éducation, elle recoupe presque parfaitement l'intensité globale de l'homogamie déjà présentée à la figure 7.3 (p. 429). La seconde dimension oppose les régions françaises aux régions italiennes, et correspond à une différence de structure de l'homogamie socioéconomique, à niveau égal.

Commençons par détailler l'opposition présente sur le premier axe. L'homogamie socioéconomique relative dans l'Est de la Roumanie (RO2, figure 7.7) est très marquée, opposant un bloc formé des classes supérieures (cadres, professions libérales et intellectuelles) aux classes populaires (ouvriers qualifiés, professions peu qualifiées), auxquelles se rattache en partie un important groupe de petits indépendants fortement endogame. Seuls les employés qualifiés et les professions intermédiaires présentent une endogamie relativement faible, et font la transition entre haut et bas de l'échelle sociale. Cependant, les effectifs de ces groupes sont assez faibles, donnant à la structure sociale de cette région-type un caractère fortement polarisé. Cette polarisation est loin d'atteindre les niveaux extrêmes déjà relevés dans le cas de l'homogamie d'éducation : les *odds ratios* généralisés les plus élevés, qui concernent l'endogamie des cadres, des professions libérales et intellectuelles et des petits indépendants, sont respectivement de 15, 18 et 27 ; ceux des autres groupes sont inférieurs à 7.

À l'opposé de cette structure sociale dans laquelle le choix du conjoint est très socialement déterminé, celle du Centre de l'Angleterre (*Midlands*, UKFG) présente des contrastes entre groupes beaucoup moins forts. Si l'opposition entre classes supérieures et

classes populaires et l'endogamie des petits indépendants sont toujours bien visibles, l'association est bien moins marquée : les *odds ratios* généralisés correspondants sont ici de 2, 3 et 3. Classes supérieures et classes populaires ne constituent pas des ensembles véritablement homogènes : si l'endogamie des groupes qui les composent reste supérieure à la moyenne, les configurations qui les croisent (cadres-professions intellectuelles supérieures, ouvriers qualifiés-professions peu qualifiées) correspondent à des *odds ratios* très peu élevés. En outre, entre ces deux pôles, on peut observer une vaste zone de choix du conjoint relativement peu marqué, que l'on peut comme dans le cas de la France (chapitre 1) identifier aux classes moyennes. Les petits indépendants ne sont plus associés aux classes populaires, mais prennent leurs conjoints de manière globalement assez peu marquée sur l'ensemble de l'échelle sociale – à l'exception d'une forte préférence pour l'endogamie.

Figure 7.6 : Dimensions du modèle à forme de régression multiple pour l'homogamie socioéconomique

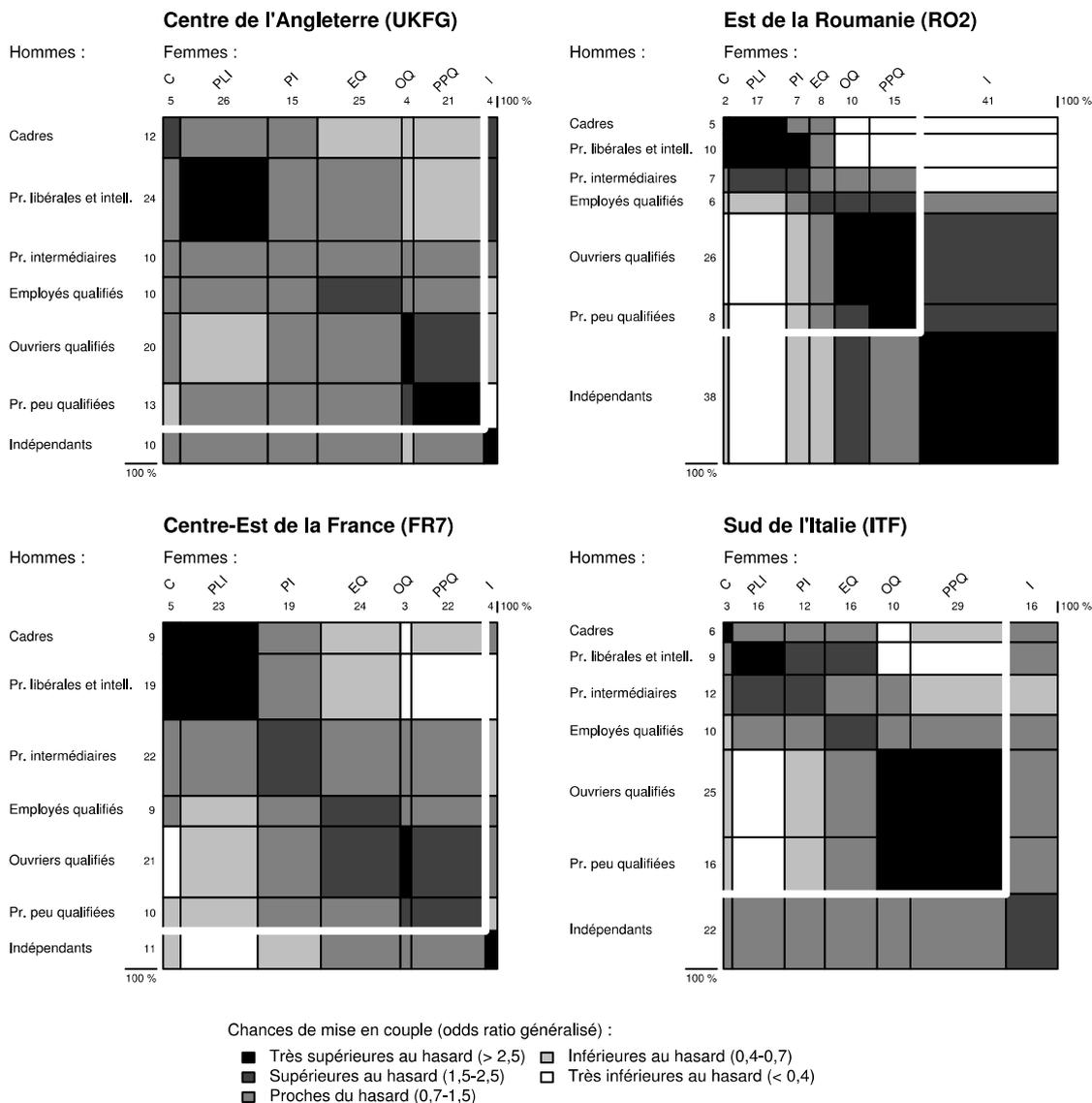


Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans et les deux conjoints ont travaillé dans les huit années précédant l'enquête

Cf. Tableau 7.5 (p. 454) pour la correspondance entre codes et noms des régions.

Figure 7.7 : Structure de l'homogamie socioéconomique dans quatre régions-typ



Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans et les deux conjoints ont travaillé dans les huit années précédant l'enquête

Dans les quatre chiasmogrammes, les hauteurs des lignes et les largeurs des colonnes correspondent aux proportions d'hommes et de femmes dans chaque catégorie parmi les personnes en couple. L'aire des cellules est égale à la proportion attendue en situation de mise en couple aléatoire, alors que la teinte de gris mesure l'écart à cette référence (*odds ratio* généralisé).

Le second axe oppose régions françaises, roumaines, allemandes et néerlandaises, aux régions italiennes, ainsi qu'à des régions espagnoles, grecques et portugaises¹⁷. On trouve dans le premier groupe des régions d'Europe occidentale (à l'exception de la Roumanie, qui constitue un cas à part, bien séparé sur le plan), et dans le second des régions du Sud de l'Europe.

La structure de l'homogamie dans le Centre-Est de la France (FR7) est relativement proche de celle du Centre de l'Angleterre (UKFG). Elle s'en distingue par une plus forte endogamie des classes supérieures, qui forment ici un bloc homogène, les *odds ratios* mesurant l'endogamie étant respectivement de 5 et 7 pour les cadres et les professions libérales et intellectuelles, et les deux configurations associant ces groupes présentant des *odds ratios* de 3. L'endogamie des classes populaires est proche dans les deux régions, mais les configurations associant des membres de cet ensemble avec des membres des classes supérieures sont légèrement moins fréquentes dans le Centre-Est de la France. Enfin, l'endogamie des petits indépendants est nettement plus forte dans cette dernière région, avec un *odds ratio* de 16 (contre 3 dans le Centre de l'Angleterre).

L'homogamie dans le Sud de l'Italie (ITF) présente des traits bien distincts. Les classes populaires (ouvriers qualifiés et professions peu qualifiées) apparaissent plus homogènes que dans le Centre-Est de la France, avec des *odds ratios* mesurant l'endogamie de 5 pour les deux groupes, et des *odds ratios* correspondant aux configurations croisant ces deux groupes de 3. Ainsi, l'endogamie des moins qualifiés est bien plus forte dans le Sud de l'Italie : l'*odds ratio* généralisé de 5 est plus de deux fois supérieur à celui mesuré pour le Centre-Est de la France (*odds ratio* de 2). Les classes supérieures présentent en revanche une endogamie de niveaux similaires dans les deux régions, avec des *odds ratios* de 6 pour les cadres comme pour les professions libérales et intellectuelles. Mais, comme dans le Centre de l'Angleterre (UKFG) et contrairement au Centre-Est de la France (FR7), ces deux groupes ne sont pas nettement associés entre eux. Enfin, l'endogamie des petits indépendants est nettement plus faible, avec un *odds ratio* de 2.

Le second axe correspond donc ici encore à une opposition dans la structure de l'homogamie plus que dans son niveau. En bas de l'axe, se situent les régions dans lesquelles

17 On peut noter comme exception en bas de l'axe Londres (UKI), et en haut de l'axe Bruxelles (BE1). Le contraste entre ces régions et celles qui les entourent est frappant.

les classes supérieures forment un ensemble homogène (mais sans que les groupes socioéconomiques qui les composent soient véritablement plus endogames) et où les petits indépendants sont fortement endogames. En haut de l'axe, ce sont les classes populaires qui présentent une forte cohérence et une forte endogamie, alors que les petits indépendants sont peu endogames. C'est paradoxalement dans ces dernières régions que la proportion d'indépendants dans la population active est la plus élevée, et la mieux équilibrée entre hommes et femmes.

Ainsi, nous observons qu'il n'existe pas de relation simple entre l'intensité de l'endogamie relative d'un groupe social et sa taille au sein d'une société. Au chapitre 2, nous avons relevé une diminution de l'endogamie des non diplômés et des agriculteurs en France depuis 1969, parallèlement à la nette diminution des effectifs de ces deux groupes. À l'inverse, J. Smits (2003) a mis en évidence une relation négative entre l'endogamie relative des diplômés du secondaire et du supérieur et la taille de ce groupe dans 55 pays. En outre, nous avons noté au chapitre 3 que l'endogamie du groupe des précaires tendait plutôt à diminuer lorsque la conjoncture économique entraînait son extension. Si la croissance numérique d'un groupe social peut ainsi être à l'origine d'une certaine dilution de son identité, cette règle n'est pas valable dans tous les cas.

5 Les variations de l'hypergamie

Les variations de l'hypergamie apparaissent elles aussi importantes entre régions d'Europe. Nous commençons par présenter celles de l'hypergamie absolue, qui reflètent principalement les inégalités entre hommes et femmes du point de vue du niveau d'éducation et de la carrière professionnelle, avant de modéliser celles de l'hypergamie relative, qui mesurent l'effet des préférences individuelles et des normes de genre sur le choix du conjoint.

5.1 L'hypergamie absolue : plus fréquente en termes socioéconomiques qu'éducatifs

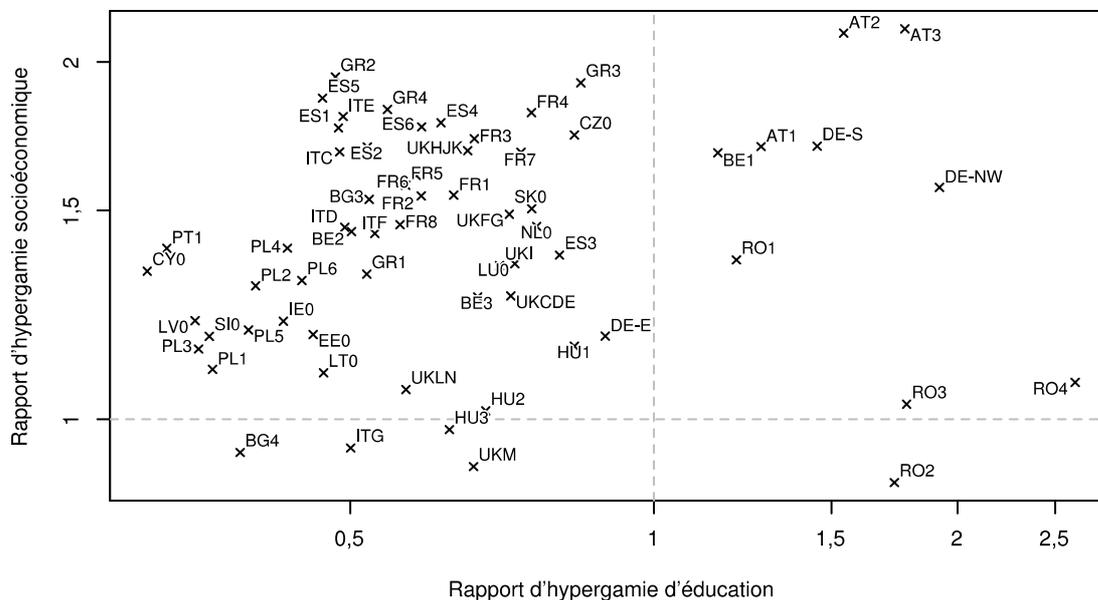
Nous mesurons l'hypergamie (ou l'hypogamie) féminine absolue à travers le rapport d'hypergamie, qui désigne le rapport entre la proportion des couples dans lesquels l'homme a un statut supérieur à celui de sa conjointe et celle des couples dans lesquels l'inverse est

vrai. Un rapport d'hypergamie supérieur à 1 indique ainsi que l'hypergamie féminine prévaut, alors qu'un rapport inférieur à 1 indique que c'est l'hypogamie féminine qui est la plus courante. Ainsi définies, les deux composantes de l'hypergamie absolue apparaissent assez modestement corrélées : le coefficient de corrélation entre les deux rapports d'hypergamie est de 0,09 seulement ; il augmente à 0,42 lorsque les régions roumaines – qui échappent ici encore aux tendances d'ensemble du fait de leur forte hypergamie d'éducation conjuguée à une faible hypergamie socioéconomique –, sont exclues de l'échantillon.

Si l'hypergamie féminine en termes socioéconomiques est bien la situation la plus fréquente dans la plupart des régions d'Europe, celle en termes d'éducation est l'exception – du moins parmi les jeunes couples (Figure 7.8). Les seules régions dans lesquelles l'hypergamie féminine d'éducation prévaut sont l'Allemagne du Sud (DE-S) et du Nord-Ouest (DE-NW) – mais pas l'Allemagne de l'Est (DE-E) –, l'Autriche, la Roumanie, ainsi que la région de Bruxelles (BE1). Ce constat général et ses exceptions sont conformes à ceux obtenus par A. Esteve, J. García-Román et I. Permanyer (2012) pour les 10 pays européens inclus dans leur échantillon à partir de données de recensement portant jusqu'en 2001, et permettent de les actualiser et de les préciser. Il apparaît que les exceptions allemande de l'Ouest et roumaine persistent en 2011.

Des variations importantes apparaissent aussi parmi les régions dans lesquelles l'hypergamie féminine d'éducation est minoritaire (moitié gauche de la figure). Ainsi, l'hypogamie féminine est particulièrement fréquente au Nord-Est de l'Europe (Pologne, Lettonie, Lituanie, Estonie), ainsi qu'à Chypre, en Bulgarie, en Slovénie, en Italie, au Portugal et en Irlande. Il est frappant de constater que cet ensemble de pays est très hétérogène, notamment du point de vue du niveau de développement économique.

Figure 7.8 : Rapport d'hypergamie d'éducation et socioéconomique dans les régions d'Europe



Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour le groupe socioéconomique)

Le rapport d'hypergamie désigne le rapport entre la proportion des couples dans lesquels l'homme a un statut supérieur à celui de sa conjointe et celle des couples dans lesquels l'inverse est vrai. Un rapport supérieur à 1 indique que l'hypergamie féminine prévaut, alors qu'un rapport inférieur à 1 indique que c'est l'hyposgamie féminine qui est la plus courante.

Une échelle logarithmique est utilisée pour permettre la comparaison directe des rapports supérieurs et inférieurs à 1, les écarts étant symétriques par rapport à l'origine.

Cf. Tableau 7.5 (p. 454) pour la correspondance entre codes et noms des régions.

De son côté, l'hypergamie socioéconomique est bien plus fréquente. Rappelons tout d'abord qu'elle ne peut par définition être mesurée que pour les couples dans lesquels les deux conjoints ont exercé une activité professionnelle. Elle peut donc offrir une mesure biaisée du statut des femmes si les régions dans lesquelles l'activité féminine est la moins fréquente présentent une faible tendance à l'hypergamie parmi les couples bi-actifs. Mais cette hypothèse n'est pas soutenue par les données : la corrélation entre le taux d'activité des femmes entre 20 et 39 ans et le rapport d'hypergamie socioéconomique est très faible, à 0,14.

Les variations de l'hypergamie socioéconomique sont assez différentes de celle en termes d'éducation. Premièrement, les régions dans lesquelles les femmes ont plus souvent un statut socioéconomique supérieur à celui de leur conjoint que l'inverse sont très peu nombreuses : la Sicile et la Sardaigne (ITG), l'Écosse (UKM), l'Est de la Roumanie (RO2) le Sud de la Bulgarie (BG4) et l'Est de la Hongrie (HU3). Lorsqu'elle a lieu, l'inversion de l'hypergamie socioéconomique reste toujours peu marquée. Parmi ces régions à hypogamie féminine, la Sicile et la Sardaigne présentent un taux d'activité féminine particulièrement faible (moins de 50 % parmi les 30-59 ans, cf. colonne « Activité féminine » dans le tableau 8.3, p. 496) : une tendance à l'hypogamie féminine parmi les couples dans lesquels les conjoints ont déjà travaillé peut masquer une éviction durable du marché du travail des femmes les moins qualifiées dans l'Europe du Sud (Mandel, 2009, p. 707). Ce lien ne semble cependant pas avoir de caractère systématique : les autres régions à hypogamie féminine présentent des taux d'activité féminine supérieurs à la moyenne européenne ; au total la corrélation entre ces deux phénomènes est très limitée (à 0,19)¹⁸.

Deuxièmement, les traces de l'action des régimes socialistes en faveur de l'activité féminine sont bien visibles (van der Lippe et Fodor, 1998 ; Brainerd, 2000), l'hypergamie étant moins marquée dans les pays de l'ancien bloc soviétique que dans la plupart des régions de l'Ouest. L'opposition entre Allemagne de l'Est (DE-E) et Allemagne de l'Ouest (DE-NW, DE-S) est très nette. En revanche, la forte hypergamie féminine de la République tchèque fait exception. En dehors des anciens pays communistes, l'hypergamie féminine en termes socioéconomiques est relativement faible en Irlande (IE0), en Irlande du Nord et au Pays de Galles (UKLN) ainsi qu'en Écosse (UKM).

5.2 L'hypergamie relative : une tendance loin d'être systématique

Les asymétries entre hommes et femmes persistent dans de nombreuses régions une fois contrôlée la structure de la population ; néanmoins, elles concernent des effectifs assez faibles, et ne sont pas décelables partout avec les données dont nous disposons.

18 Nous introduirons au chapitre suivant une mesure des inégalités entre hommes et femmes en termes de statut professionnel qui résout cette difficulté en considérant l'inactivité comme un statut professionnel inférieur à tous les autres. Toutefois, la mesure de l'hypergamie telle que définie ici garde son intérêt comme approche du statut social des conjoints.

Concernant l'hypergamie d'éducation, nous partons du modèle d'indépendance conditionnelle (Tableau 7.4). Si celui-ci contient uniquement des paramètres permettant de contrôler la structure des populations masculine et féminine dans chaque région, il est dans ce cas particulier équivalent à un modèle quasi-symétrique du fait de l'exclusion de la diagonale de la table et du faible nombre de catégories retenues (*cf.* p. 422). Il décrit déjà très précisément les données, en classant correctement plus de 98 % des couples : on voit que l'ampleur des asymétries en termes relatifs est très limitée. Un second modèle postule qu'une tendance à l'hypergamie ou à l'hypogamie féminine gouverne le choix du conjoint au-delà des contraintes structurelles, mais que cette tendance est la même dans toutes les régions d'Europe. Cette hypothèse est clairement rejetée : ce modèle n'améliore absolument pas la description des données selon la déviance et l'indice de dissimilarité, et les indicateurs BIC et AIC augmentent. Ce résultat ne revient cependant pas à affirmer que les asymétries hommes-femmes n'existent dans aucune région d'Europe. En effet, le troisième modèle, autorisant l'hypergamie (ou l'hypogamie), à varier librement selon les régions, améliore la description des données. Ce troisième modèle étant équivalent au modèle saturé, il reconstruit parfaitement les tables d'homogamie. Ce ne sont donc pas la déviance et l'indice de dissimilarité, nuls par définition, qui doivent retenir notre attention, mais le BIC et l'AIC : si le premier préfère le modèle d'indépendance, le modèle saturé a les faveurs du second, signe de l'existence d'asymétries hommes-femmes statistiquement significatives au moins dans certaines régions.

L'hypergamie est plus nette – du moins en termes d'effectifs concernés – du point de vue du groupe socioéconomique, ce qui tient en partie à la classification plus détaillée que nous utilisons. Ici, le modèle d'indépendance est distinct du modèle de quasi-symétrie et du modèle quasi-symétrique à forme de régression à deux dimensions. Ce dernier modèle, similaire à celui retenu pour la description de l'homogamie dans la section précédente, nous sert de référence pour décrire précisément la composante symétrique de l'association et nous assurer qu'elle n'interfère pas avec la mesure de l'hypergamie. On voit qu'il ne décrit pas aussi fidèlement les données que dans le cas de l'éducation : près de 7 % des couples sont classés dans une mauvaise cellule. Le modèle postulant une hypergamie stable, s'il ne permet pas de diminuer ce chiffre de manière nette, est néanmoins préféré par le BIC et l'AIC, indiquant qu'il existe bien une tendance à l'asymétrie hommes-femmes qui va dans le même sens dans l'ensemble de l'Union européenne – ce qui n'était pas le cas pour

l'éducation. Mais s'en tenir à ce résumé serait simplificateur : le dernier modèle, autorisant ces asymétries à changer de sens et d'ampleur entre régions, permet de classer correctement 0,5 % des couples supplémentaires, et se voit préféré par l'AIC, sinon par le BIC.

Tableau 7.4 : Indicateurs d'ajustement des modèles concernant l'hypergamie

	D. L.	Déviante	Δ (%)	BIC	AIC
Éducation :					
Indépendance conditionnelle	64	255	1,44	-472	127
Stabilité de l'hypergamie	63	255	1,46	-460	129
Variation de l'hypergamie	0	0	0,00	0	0
Groupe socioéconomique :					
Indépendance conditionnelle	1 856	19 446	14,87	-2 338	15 734
Quasi-symétrie stable	1 842	7 381	8,23	-14 239	3 697
Quasi-sym. à forme de régression, 2 dimensions	1 692	5 702	6,72	-14 157	2 318
Stabilité de l'hypergamie	1 691	5 684	6,70	-14 164	2 302
Variation de l'hypergamie	1 567	5 113	6,23	-13 279	1 979

D. L. : degrés de liberté. Δ : indice de dissimilarité.

5.2.1 La significativité statistique de l'hypergamie relative

Dans le cas de l'éducation comme du groupe socioéconomique, il apparaît donc que des asymétries entre hommes et femmes sont décelables, et qu'elles varient entre régions d'Europe. Celles-ci n'ont cependant rien de massif, comme l'indiquent les évolutions de l'indice de dissimilarité : les variations de l'hypergamie absolue, déjà décrites plus haut, constituent le phénomène principal du point de vue des inégalités hommes-femmes. En outre, si le modèle autorisant des variations entre régions a les faveurs de l'AIC, le coefficient d'hypergamie n'est pas significatif dans nombre des régions étudiées. Nous ne présentons donc dans les cartes ci-dessous (Figure 7.9) que les régions pour lesquelles la tendance à l'hypergamie ou à l'hypogamie est significative au seuil de 10 %¹⁹.

L'absence de significativité des asymétries hommes-femmes dans une région donnée peut correspondre à deux situations différentes. D'une part, elle peut indiquer l'absence

¹⁹ Ces tendances sont aussi significatives au seuil de 5 % dans 13 régions pour l'éducation, et dans 25 régions pour le groupe socioéconomique.

d'asymétries autres que d'ampleur négligeable. D'autre part, elle peut découler d'une taille d'échantillon trop faible pour la détection d'un effet relativement exigeant en termes de puissance statistique. En effet, une proportion importante des couples étant endogames, l'étude des asymétries hommes-femmes repose nécessairement sur un échantillon assez restreint des couples²⁰. Comme on l'a vu ci-dessus, une fois la structure de la population contrôlée, les écarts à la symétrie concernent une proportion réduite des effectifs.

Néanmoins, il semble que les difficultés liées à la taille des échantillons n'expliquent pas complètement l'absence de significativité statistique. Les intervalles de confiance autour des valeurs estimées pour le paramètre d'hypergamie sont raisonnablement étroits dans une proportion importante des cas. En ce qui concerne l'hypergamie d'éducation, un intervalle de confiance à 90 % permet de circonscrire entre 0,5 et 2 la valeur de l'*odds ratio* généralisé anti-symétrique de 6 régions sur les 22 pour lesquelles le coefficient d'hypergamie n'est pas significativement différent de zéro (et à 95 % pour 5 régions)²¹. Concernant l'hypergamie socioéconomique, parmi les 21 régions pour lesquelles ce paramètre n'est pas significatif, les intervalles de confiance à 90 % nous permettent même de circonscrire sa valeur entre 0,8 et 1,25 pour 11 d'entre elles (et à 95 % pour 7 d'entre elles). Ainsi, il apparaît que l'absence de significativité reflète pour une part la faiblesse réelle des asymétries, et pas seulement des difficultés liées à la taille des échantillons régionaux utilisés.

Finalement, notons que lorsque l'unité d'analyse retenue est le pays plutôt que la région – ce qui augmente nettement le rapport entre effectifs et nombre de paramètres à estimer –, les résultats obtenus sont très similaires : les pays dans lesquels les asymétries sont significatives sont ceux qui comprennent des régions dans le même cas²². Au total, s'il serait évidemment utile de reproduire ces analyses à partir d'échantillons plus importants de

20 Ce problème est aggravé par le recours à une classification très agrégée du niveau d'éducation : moins ce niveau est détaillé, et plus on aura tendance à considérer deux conjoints comme de même niveau, quand une distinction plus fine aurait fait apparaître un écart.

21 Dans le cas de la Slovaquie, dont la situation apparaissait déjà très particulière du point de vue de l'homogamie d'éducation, l'intervalle de confiance est extrêmement large, et l'*odds ratio* généralisé anti-symétrique estimé supérieur à 1000. Dans ce qui suit, nous ramenons cette valeur au niveau de la seconde région à la plus forte hypergamie, le Sud de la Roumanie (RO3), dont l'*odds ratio* généralisé anti-symétrique est égal à 17.

22 Du point de vue de l'éducation, les asymétries sont significatives au seuil de 10 % aux Pays-Bas, au Royaume-Uni et en Bulgarie (hypogamie féminine), ainsi qu'en Allemagne (Est et Ouest séparément), en République tchèque, et en Roumanie (hypergamie féminine). Du point de vue du groupe socioéconomique, c'est le cas au Royaume-Uni (hypogamie féminine), ainsi qu'en Bulgarie, Italie, France, Grèce, Lettonie, Lituanie, Pologne, Portugal (seule exception à la règle), République tchèque et Roumanie (hypergamie féminine).

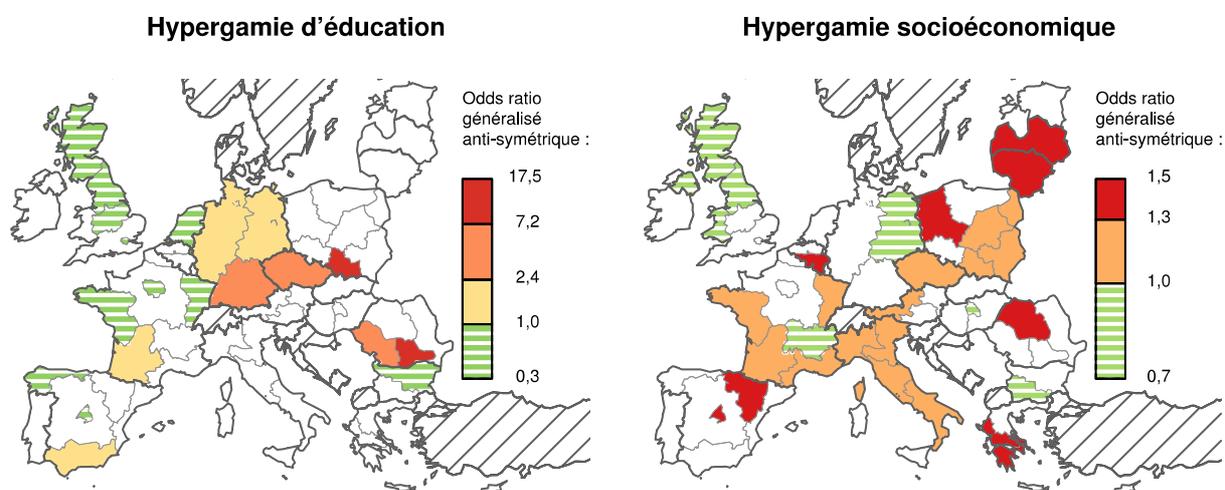
manière à préciser la mesure de l'hypergamie en Europe, les résultats semblent assez fiables pour servir de base à l'étude des déterminants de ce phénomène, qui sera réalisée au chapitre suivant.

5.2.2 La géographie de l'hypergamie relative en Europe

Les *odds ratios* généralisés anti-symétriques calculés à partir du modèle autorisant l'hypergamie à varier librement entre régions font apparaître que ce phénomène n'a rien de généralisé en Europe (Figure 7.9 ; les valeurs exactes sont reprises dans le tableau 8.3, p. 496). Dans plusieurs régions, c'est l'hypogamie féminine relative qui prévaut, aussi bien du point de vue du niveau d'éducation des conjoints que de leur groupe socioéconomique. Ainsi, l'hypogamie féminine d'éducation domine dans une large moitié Nord de la Grande-Bretagne et à Londres, aux Pays-Bas, dans certaines régions du Nord de la France, dans le Nord-Est de l'Espagne et à Madrid, ainsi que dans le Nord-Est de la Bulgarie. L'hypogamie féminine en termes socioéconomiques domine quant à elle dans le Nord du Royaume-Uni, au Centre-Est de la France, en Allemagne de l'Est, dans la région de Budapest et dans le Sud-Ouest de la Bulgarie.

À l'inverse, l'hypergamie féminine d'éducation l'emporte dans toutes les régions d'Allemagne, en République tchèque, dans le Sud de la Pologne et de la Roumanie, ainsi que dans le Sud-Ouest de la France et dans le Sud de l'Espagne. Du point de vue socioéconomique, elle domine dans une grande partie de la France et de l'Italie, en Wallonie, en Lettonie, en Lituanie, dans la plus grande partie de la Pologne, en République tchèque, ainsi que dans certaines régions d'Autriche, de Roumanie, de Grèce et d'Espagne.

Figure 7.9 : Géographie de l'intensité de l'hypergamie relative d'éducation et socioéconomique en Europe



Source : Enquête européenne sur les forces de travail 2011 (Eurostat)

Champ : couples cohabitants dans lesquels l'un des deux conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour l'hypergamie socioéconomique)

Les régions non colorées sont celles pour lesquelles aucune tendance à l'hypergamie ou à l'hypogamie statistiquement significative n'a pu être mesurée. Une échelle logarithmique est utilisée pour permettre la comparaison directe des valeurs supérieures et inférieures à 1, les écarts étant symétriques par rapport à cette référence. Le découpage en classes a été réalisé automatiquement à l'aide de l'algorithme de Jenks (1967), en introduisant une borne obligatoire à 1.

L'*odds ratio* généralisé anti-symétrique (Encadré 4.1, p. 259) est égal au rapport entre les *odds ratios* généralisés correspondant aux configurations à hypergamie féminine et ceux correspondant aux configurations à hypogamie féminine. Une valeur supérieure à 1 indique une tendance à l'hypergamie féminine au-delà de ce qu'impose la structure de la population, alors qu'une valeur inférieure à 1 indique une tendance à l'hypogamie féminine.

Le tableau incomplet de l'Europe que dessinent les paramètres significatifs laissent apparaître des convergences et des divergences nationales. Ainsi, le Royaume-Uni et la Bulgarie semblent constituer deux pôles d'hypogamie féminine assez homogènes du point de vue des deux composantes étudiées. À l'inverse, la République tchèque, la Pologne et la Roumanie comprennent des régions à forte hypergamie – quoique la significativité varie selon la composante retenue.

Certains pays présentent en revanche des divergences importantes entre hypergamie d'éducation et hypergamie socioéconomique. L'Allemagne présente une très nette tendance à l'hypergamie d'éducation, mais pas à l'hypergamie socioéconomique – l'Allemagne de l'Est constituant même l'une des rares zones d'hypogamie en Europe centrale. De même, l'hypergamie socioéconomique prévaut clairement dans la plus grande partie de l'Italie, mais aucune tendance à l'hypergamie d'éducation n'y est décelable. Enfin, d'autres pays sont divisés pour une même composante, comme c'est le cas en France pour les deux composantes, ou en Espagne pour l'éducation.

Quelles régularités peut-on identifier dans ces variations ? Les deux composantes de l'hypergamie relative ne sont absolument pas liées entre elles : que l'on retienne les 64 régions ou que l'on se limite aux 9 dans lesquelles les asymétries hommes-femmes sont significatives selon les deux composantes, le coefficient de corrélation est inférieur à 0,15. Il semble donc difficile de faire de l'hypergamie relative en termes d'éducation et de groupe socioéconomique deux indicateurs alternatifs du statut des femmes dans une société.

En revanche, chaque composante de l'hypergamie relative est corrélée de manière non négligeable à sa mesure absolue. Ainsi, confirmant nos observations des chapitres 4 et 5 concernant l'inversion de l'hypergamie en France, la supériorité scolaire des femmes semble être liée de manière relativement nette à l'hypergamie relative : la corrélation entre hypergamie absolue et hypergamie relative d'éducation est de 0,39, et de 0,57 lorsque l'on s'en tient aux régions dans lesquelles les asymétries hommes-femmes en termes d'éducation sont significatives. Parmi ces dernières, seules 5 présentent une discordance entre hypergamie absolue et hypergamie relative : en Allemagne de l'Est, en République tchèque, dans le Sud de la Pologne, dans le Sud-Ouest de la France et dans le Sud de l'Espagne, on note une tendance à l'hypergamie féminine relative conjuguée à une hypogamie absolue ; aucun cas inverse (hypogamie relative conjuguée à une hypergamie absolue) n'est à relever. Nous pouvons faire l'hypothèse que ces régions sont observées au milieu de la transition vers l'hypogamie féminine observée au niveau mondial (Esteve, García-Román et Permanyer, 2012), dans la fenêtre relativement courte précédant l'évolution des normes orientant le choix du conjoint. Il resterait à déterminer si cet écart entre structures et comportements est le signe d'une augmentation récente du niveau d'éducation des femmes, ou celui d'une plus forte résistance des normes de genre, dans les régions considérées. Nous

ne pouvons guère approfondir cette piste à partir des données purement transversales dont nous disposons.

De son côté, la corrélation entre hypergamie socioéconomique absolue et relative n'est que de 0,24, mais elle s'élève à 0,48 lorsque seules les régions à asymétries significatives sont retenues. Le croisement entre les deux mesures ne permet pas d'évaluer si le même phénomène est à l'œuvre que pour l'hypergamie d'éducation : trop peu de régions présentent une hypergamie relative statistiquement significative du point de vue du groupe socioéconomique.

6 Conclusion

Au terme de cette première étude descriptive, nous avons mis en évidence des différences très nettes tant dans le niveau que dans la structure de l'homogamie et de l'hypergamie entre les régions d'Europe. Ce résultat confirme l'intérêt de retenir ce champ géographique pour l'analyse des déterminants du choix du conjoint que nous réaliserons au chapitre suivant.

Nous avons tout d'abord observé une corrélation relativement forte entre homogamie d'éducation et socioéconomique, aussi bien en termes absolus que relatifs (coefficients de 0,6). Pour chacune de ces dimensions, nous avons en outre mesuré une corrélation elle aussi importante entre homogamie absolue et homogamie relative (coefficients entre 0,4 et 0,9). Cependant, l'homogamie d'éducation apparaît nettement plus élevée que l'homogamie socioéconomique dans la quasi-totalité des régions. Sa distribution géographique est aussi la plus nette : elle décroît d'Est en Ouest, avec quelques exceptions comme les États baltes et la Grèce (faible homogamie) ou le Portugal (forte homogamie) ; l'homogamie la plus forte est observée en Roumanie et en Bulgarie. L'homogamie socioéconomique se répartit de manière plus complexe : si elle est plus forte dans l'Est et le Sud de l'Europe, cette régularité n'est que très partielle ; c'est d'abord dans les régions à dominante catholique qu'elle apparaît la plus élevée, avec comme exceptions les plus notables l'Irlande (faible homogamie) et la Roumanie (forte homogamie). Enfin, nous avons observé une homogamie socioéconomique légèrement plus forte dans les grandes capitales que dans les régions qui les entourent.

Des différences dans la structure de l'homogamie sont elles aussi apparues. Du point de vue de l'éducation, une dimension secondaire oppose les régions dans lesquelles l'endogamie est forte au milieu et au sommet de l'échelle des diplômes, à ceux dans lesquels elle est plus forte parmi les moins éduqués. Nous avons vu que cette opposition recoupait en grande partie des différences dans l'organisation des systèmes éducatifs, et plus précisément dans la position des paliers de sélection principaux. Du point de vue du groupe socioéconomique, la seconde dimension oppose les régions d'Europe du Sud à certaines régions d'Europe de l'Ouest. Les premières se caractérisent par une faible cohésion des classes supérieures, divisées entre cadres et professions libérales et intellectuelles, qui contraste avec une forte cohésion des classes populaires, ainsi que par une endogamie des petits indépendants relativement faible. Au contraire, dans les secondes, ce sont les classes supérieures qui sont homogènes et les classes populaires plus divisées entre qualifiés et non qualifiés, alors que les petits indépendants sont fortement endogames.

Les variations de l'hypergamie sont elles aussi sensibles en termes absolus comme relatifs. Cependant, l'hypergamie relative apparaît comme un déterminant très secondaire du choix du conjoint : sa prise en compte n'améliore que très légèrement la description des tables, et les données ne nous permettent pas de mettre en évidence des tendances statistiquement significatives dans toutes les régions. Ainsi, les inégalités de statut entre hommes et femmes au sein du couple sont très largement dues aux inégalités genrées au sein de la population. Néanmoins, ni l'hypergamie absolue ni l'hypergamie relative ne sont absolument systématiques en Europe : du point de vue de l'éducation comme du groupe socioéconomique, des tendances à l'hypergamie comme à l'hypogamie féminine peuvent être observées dans plusieurs régions, en termes relatifs aussi bien qu'absolus.

Si hypergamie d'éducation et socioéconomique sont corrélées de manière modeste en termes absolus, avec un coefficient de 0,4 (une fois exclues les régions roumaines qui échappent au schéma d'ensemble), elles ne présentent aucun lien en termes relatifs. Il apparaît donc qu'hypergamie socioéconomique et hypergamie d'éducation ne renvoient pas, dans leur mesure relative, au même phénomène social. La faible précision de nos estimations (qui découle notamment du manque de détail de la classification du niveau d'éducation utilisée), si elle joue sans doute un rôle, ne semble pas pouvoir expliquer à elle seule la faiblesse de cette corrélation. De fait, la mesure absolue et la mesure relative sont

corrélées pour chaque composante de l'hypergamie (coefficients entre 0,5 et 0,6), ce qui paraît indiquer que l'hypergamie relative est correctement appréhendée, et qu'elle est liée au niveau des inégalités entre hommes et femmes du point de vue éducatif ou socioéconomique.

Ces deux constats – corrélation des mesures absolue et relative sur une même dimension, mais discordance sur des dimensions différentes – ne sont pas contradictoires. Ils nous amènent à conclure que les sociétés diffèrent par les dimensions de la stratification sociale selon lesquelles ces inégalités s'expriment, une égalité – voire une supériorité – des femmes dans un domaine pouvant très bien se combiner à de fortes inégalités dans un autre. Nous confirmerons cette interprétation au chapitre suivant.

Il se pourrait aussi que la faiblesse du lien mesuré pour une seule année entre ces deux phénomènes provienne de l'opposition entre les évolutions temporelles qui sont à l'œuvre. On sait que l'hypergamie féminine d'éducation s'est inversée dans de nombreux pays, d'abord en termes relatifs, puis très rapidement après en termes absolus. Au contraire, nous avons observé au chapitre 4 dans le cas de la France que l'hypergamie socioéconomique s'était d'abord effondrée en termes relatifs, les inégalités de carrières entre hommes et femmes opposant une plus forte inertie. Cette hypothèse ne pourrait être creusée qu'à l'aide de données portant sur plusieurs années dans plusieurs pays européens.

7 Appendice : Détail des régions étudiées

Tableau 7.5 : Régions retenues et effectifs correspondants dans l'échantillon

Code	Pays ou région(s)	Effectif éducation	Effectif ESeG
AT	Autriche	13 057	11 234
AT1	Ostösterreich	4 244	3 611
AT2	Südösterreich	2 635	2 304
AT3	Westösterreich	6 178	5 319
BE	Belgique	5 218	4 202
BE1	Région de Bruxelles-Capitale	740	409
BE2	Vlaams Gewest	2 766	2 431
BE3	Région wallonne	1 712	1 362

Code	Pays ou région(s)	Effectif éducation	Effectif ESeG
BG	Bulgarie	1 846	1 412
BG3	Severna i yugoiztochna Bulgaria	1 013	730
BG4	Yugozapadna i yuzhna tsentralna Bulgaria	833	682
CZ0	République tchèque	3 538	2 937
CY0	Chypre	3 579	3 032
DE	Allemagne	3 051	10 015
DE-S	Allemagne du Sud : Baden-Württemberg (DE1)/Bayern (DE2)	605	493
DE-E	Allemagne de l'Est et Berlin Ouest : Berlin (DE3)/Brandenburg (DE4)/Mecklenburg-Vorpommern (DE8)/Sachsen (DED)/Sachsen-Anhalt(DEE)/Thüringen (DEG)	1 483	1 152
DE-NW	Allemagne du Nord-Ouest : Hamburg (DE6)/Hessen (DE7)/Niedersachsen (DE9)/Nordrhein-Westfalen (DEA)/Rheinland-Pfalz (DEB)/Saarland (DEC)/Schleswig-Holstein (DEF)	963	798
EE0	Estonie	1 737	1 462
ES	Espagne	6 501	5 302
ES1	Noroeste	1 002	808
ES2	Noreste	902	764
ES3	Comunidad de Madrid	343	293
ES4	Centro	1 195	978
ES5	Este	1 530	1 273
ES6	Sur	1 529	1 186
FR	France	12 785	8 462
FR1	Île-de-France	2 338	1 543
FR2	Bassin Parisien	2 508	1 627
FR3	Nord - Pas-de-Calais	1 027	614
FR4	Est	1 205	799
FR5	Ouest	1 746	1 289
FR6	Sud-Ouest	1 339	922
FR7	Centre-Est	1 366	892
FR8	Méditerranée	1 256	776

Code	Pays ou région(s)	Effectif éducation	Effectif ESeG
GR	Grèce	16 827	10 780
GR1	Voreia Ellada	5 304	3 341
GR2	Kentriki Ellada	4 685	2 806
GR3	Attiki	4 369	3 049
GR4	Nisia Aigaiou, Kriti	2 469	1 584
HU	Hongrie	19 813	14 559
HU1	Közép-Magyarország	3 594	2 890
HU2	Dunántúl	6 621	5 165
HU3	Alföld és Észak	9 598	6 504
IE0	Irlande	17 742	11 160
IT	Italie	36 697	24 404
ITC	Nord-Ovest	9 735	7 569
ITD	Nord-Est	7 772	5 981
ITE	Centro	5 731	4 197
ITF	Sud	8 898	4 527
ITG	Isole	4 561	2 130
LT0	Lituanie	3 589	3 033
LU0	Luxembourg	1 632	1 380
LV0	Lettonie	2 280	1 876
NL0	Pays-Bas	7 125	5 858
PL	Pologne	30 775	24 090
PL1	Region Centralny	4 745	3 936
PL2	Region Poludniowy	4 319	3 338
PL3	Region Wschodni	7 381	5 793
PL4	Region Północno-Zachodni	5 328	4 154
PL5	Region Poludniowo-Zachodni	3 685	2 862
PL6	Region Północny	5 317	4 007

Code	Pays ou région(s)	Effectif éducation	Effectif ESeG
PT	Portugal	7 189	6 579
PT1	Continente	7 189	6 579
RO	Roumanie	17 357	12 391
RO1	Macroregiunea unu (Nord-Ouest et Centre)	4 279	2 996
RO2	Macroregiunea doi (Est)	5 196	3 679
RO3	Macroregiunea trei (Sud)	4 434	3 378
RO4	Macroregiunea patru (Ouest)	3 448	2 338
SI0	Slovénie	3 507	2 912
SK0	Slovaquie	6 721	4 444
UK	Royaume-Uni	6448	5 428
UKCDE	Nord de l'Angleterre : North East (UKC)/North West (UKD)/Yorkshire and The Humber (UKE)	1 602	1 328
UKFG	Midlands : East Midlands (UKF)/West Midlands (UKG)	1 088	933
UKHJK	Sud et Est de l'Angleterre : East of England (UKH)/South East (UKJ)/South West (UKK)	1 954	1 685
UKI	London	777	574
UKLN	Pays de Galles (UKL) et Irlande du Nord (UKN)	502	454
UKM	Scotland	525	454

L'effectif désigne le nombre de couples finalement retenus après application des limites d'âge et élimination des valeurs manquantes.

CHAPITRE 8

LES FACTEURS DE L'HOMOGAMIE ET DE L'HYPERGAMIE SOCIOÉCONOMIQUE ET D'ÉDUCATION : OUVERTURE SOCIALE ET LIBÉRALISME CULTUREL PLUS QUE NIVEAU DE DÉVELOPPEMENT

Après avoir mis en évidence au chapitre précédent des variations notables de l'homogamie et de l'hypergamie en termes d'éducation et de groupe socioéconomique entre régions d'Europe, nous concluons cette étude par une analyse des déterminants de ces deux phénomènes. Notre objet sera ainsi de mettre à l'épreuve le cadre théorique développé tout au long de cette thèse, et notamment les théories de la modernisation, dont nous avons vu que la littérature fournissait des évaluations relativement divergentes.

Il s'agit ici de conduire une analyse au niveau macrosocial (celui des régions) plutôt qu'à celui des individus : nous nous intéressons ici aux déterminants de la structure sociale. Nous porterons en outre exclusivement notre attention sur les mesures relatives de l'homogamie et de l'hypergamie, ce qui permet d'évacuer directement l'influence mécanique de la structure de la population de chaque région, pour nous concentrer sur des mécanismes moins évidents et sociologiquement plus pertinents.

Nos données étant purement transversales, nous ne chercherons pas à traiter de la question des évolutions temporelles du choix du conjoint prévues par les théories de la modernisation : seuls les mécanismes qu'elles prévoient seront étudiés. Tenter d'inférer à partir de résultats portant sur l'année 2011 un schéma d'évolution que devraient suivre l'ensemble des sociétés européennes est en effet un exercice périlleux dont les limites sont bien connues (Thornton, 2001). On sait que la question d'une convergence ou d'une persistance des différences de modèles familiaux en Europe fait toujours l'objet de débats. Si

les études comparatives à la fois dans l'espace et le temps, dont la littérature offre plusieurs exemples (Ultee et Luijkx, 1990 ; Smits, Ultee et Lammers, 1999, 2000 ; Smits et Park, 2009 ; Monaghan, 2015), présentent des avantages certains par rapport aux études transversales, un tel objectif exigerait de renoncer à l'apport principal de notre approche : la qualité des données, la variété et la précision des indicateurs disponibles sur le champ restreint de l'Union européenne des années récentes. Du fait de la forte inertie des phénomènes étudiés, une étude longitudinale n'a d'intérêt que si les points temporels étudiés sont séparés de plusieurs décennies, ce qui nous ramènerait à des périodes dans lesquelles la disponibilité des données est problématique.

L'homogamie et l'hypergamie relatives que nous chercherons ici à expliquer sont reprises des mesures du chapitre précédent fondées sur l'Enquête européenne sur les forces de travail 2011. Nous commençons ce chapitre en présentant une nouvelle source de données, l'Enquête sociale européenne (*European Social Survey*, ESS), à partir de laquelle nous calculerons une partie des indicateurs concernant les déterminants potentiels du choix du conjoint. Puis nous récapitulons les hypothèses qui nous semblent devoir être testées, d'abord concernant l'homogamie, puis concernant l'hypergamie. Ces hypothèses découlent des théories de la modernisation, enrichies par notre discussion des facteurs de l'hypergamie et des résultats de la littérature présentée au chapitre précédent. Pour chacune, nous présentons un ou plusieurs indicateurs qui permettent de les opérationnaliser.

Nous proposons ensuite de recourir à une technique peu utilisée en sociologie, la régression PLS, capable de prendre en charge la modélisation conjointe de quatre variables (l'homogamie et l'hypergamie en termes d'éducation et de groupe socioéconomique) à partir d'un très grand nombre de variables explicatives (28 indicateurs au total), alors que le nombre d'observations est relativement réduit (64 régions). Nous tenterons de montrer que cette technique permet de répondre à des difficultés théoriques et méthodologiques classiques des analyses comparatives macrosociologiques.

Les résultats confirment largement nos hypothèses théoriques. Malgré des divergences d'ordre secondaire, les déterminants de l'homogamie et de l'hypergamie apparaissent fondamentalement communs, aussi bien en termes d'éducation que de groupe socioéconomique. Le degré d'ouverture sociale et de libéralisme culturel sont les principaux facteurs d'une faible détermination sociale du choix du conjoint. Si la massification scolaire

et l'activité féminine sont négativement associées à l'homogamie et à l'hypergamie, le rôle des inégalités sociales est plus ambigu.

1 Données : l'Enquête sociale européenne et d'autres sources

L'homogamie et l'hypergamie en termes relatifs sont reprises des modèles estimés au chapitre précédent concernant les couples cohabitants (mariés ou non) dans lesquels l'un des conjoints est âgé de 20 à 39 ans, sur les données de l'Enquête européenne sur les forces de travail 2011. L'homogamie est mesurée par le logarithme¹ de l'*odds ratio* absolu moyen (MAOR) calculé au chapitre précédent à partir du modèle à forme de régression à deux dimensions. L'hypergamie est appréhendée à travers le logarithme de l'*odds ratio* généralisé anti-symétrique calculé au chapitre précédent à partir du paramètre d'hypergamie spécifique à chaque région.

Les variables explicatives sont mesurées dans les années 2000, les sources de données ne permettant pas de remonter au-delà pour l'ensemble des pays considérés. La mesure de l'homogamie et de l'hypergamie étant réalisée à partir des couples de 20 à 39 ans, dont la formation est presque certainement postérieure à 1990 (les conjoints les plus âgés ayant 18 ans à cette date), prendre cette période pour référence semble raisonnable, dans la mesure où les phénomènes étudiés présentent une grande inertie. Le champ retenu correspond à l'ensemble des personnes âgées de plus de 15 ans (sauf exceptions pour des raisons théoriques, signalées ci-dessous). En effet, on fait l'hypothèse que l'ensemble de la société, et pas seulement les cohortes concernées, influencent la mise en couple. En outre, les transformations sociales ayant une forte composante générationnelle, retenir les cohortes précédentes permet de se rapprocher du contexte qui prévalait lors de la formation (dans les années 1990) des couples les plus anciens.

Les indicateurs proviennent de plusieurs sources, que nous précisons ci-dessous lors de leur description détaillée. Certains indicateurs officiels sont repris des publications d'Eurostat ou d'autres bases de données internationales. Lorsque nécessaire, d'autres sont calculés à partir de l'Enquête européenne sur les forces de travail 2011. Enfin, les variables

1 Le modèle de régression PLS étant linéaire, et l'*odds ratio* étant exprimé sur une échelle multiplicative, le passage par le logarithme s'impose. Les effets estimés dans le modèle correspondent donc à des multiplications des *odds ratios*.

portant sur les déterminants culturels sont tirées des vagues 1 à 5 de l'Enquête sociale européenne (*European Social Survey*)², collectées entre 2002 et 2010. Au total, 229 000 individus ont été interrogés. Sur 23 pays, 11 n'ont pas participé à l'ensemble des vagues d'enquêtes (Tableau 8.1) ; cependant, tous ont participé à au moins deux vagues différentes, et la plupart à trois vagues. Les régions NUTS1 ont été harmonisées entre vagues d'enquête à partir d'une table de correspondance existante (Róžańska-Putek *et al.*, 2009), avant d'être fusionnées pour correspondre aux regroupements retenus au chapitre précédent. Les effectifs par régions sont presque toujours supérieurs à 500 (à l'exception de deux régions), et supérieurs à 1000 pour trois quarts des régions. Les pondérations tenant compte du plan d'enquête retenu par chaque pays sont systématiquement utilisées.

On trouvera une liste complète des valeurs des variables explicatives et expliquées pour chaque région, ainsi que des effectifs de l'Enquête sociale européenne, dans le tableau 8.3 (p. 496) en appendice à ce chapitre. Afin de permettre au lecteur d'évaluer les liens entre ces variables, leurs corrélations deux à deux sont représentées graphiquement dans la figure 8.3 (p. 495).

2 La majorité des fichiers ont été générés à partir des bases cumulatives. Pour quatre pays (Autriche, Lettonie, Lituanie et Roumanie), les données séparées des vagues 3 et/ou 4, qui ne font pas partie des bases cumulatives du fait de l'absence de poids ou d'un léger décalage temporel dans la passation de l'enquête, ont été intégrées séparément. Les références exactes des bases de données utilisées sont les suivantes : ESS Round 1: European Social Survey Round 1 Data (2002). Data file edition 6.3. ESS Round 2: European Social Survey Round 2 Data (2004). Data file edition 3.3. ESS Round 3: European Social Survey Round 3 Data (2006). Data file edition 3.4. ESS Round 4: European Social Survey Round 4 Data (2008). Data file edition 4.1. Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data.

Tableau 8.1 : Vagues de l'Enquête sociale européenne auxquelles ont participé les différents pays

	Vague 1 (2002)	Vague 2 (2004)	Vague 3 (2006)	Vague 4 (2008)	Vague 5 (2010)
Bulgarie			×	×	×
Chypre			×	×	×
République tchèque	×	×		×	×
Estonie		×	×	×	×
Grèce	×	×		×	×
Italie	×	×			
Lettonie			×	×	
Lituanie				×	×
Luxembourg	×	×			
Roumanie			×	×	
Slovaquie		×	×	×	×

Les pays ayant participé à toutes les vagues ne sont pas repris ici. On pourra se référer au chapitre précédent pour la liste complète des 23 pays étudiés.

2 Les déterminants de l'homogamie : des hypothèses aux variables

Dans cette section, nous récapitulons les hypothèses à tester au regard du cadre théorique développé tout au long de cette thèse, de la littérature comparative existante (*cf.* chapitre précédent) et des résultats déjà établis dans le cas de la France, en les accompagnant à chaque fois des variables qui nous permettront de les opérationnaliser. Nous distinguons sept grands ensembles de facteurs : opportunités de rencontres, entrecroisement des appartenances, inégalités sociales, niveau de vie, libéralisme des mœurs et contrôle social, statut des femmes et religion.

2.1 Opportunités de rencontre

Un premier ensemble de déterminants de l'homogamie a trait aux opportunités de rencontre. L'importance de cette dimension pour le choix du conjoint a notamment été démontrée par des études portant sur les lieux de rencontre (Bozon et Héran, 1987, 1988 ; Kalmijn et Flap, 2001 ; Bozon et Rault, 2012). Deux tendances majeures, de sens opposés, peuvent être anticipées.

Premièrement, le milieu agricole est généralement le moins favorable aux rencontres hétérogames, du fait d'un isolement géographique et d'une forte homogénéité sociale au niveau local. L'exode rural, en affaiblissant l'identité du monde agricole et en limitant les opportunités de rencontre proches, peut entraîner une diminution de l'endogamie – à la fois en termes socioéconomiques et d'éducation³. Ce phénomène a déjà été observé aux chapitres 1 et 2 dans le cas de la France, où il expliquait une part importante de l'affaiblissement de l'homogamie tant absolue que relative au cours des quarante dernières années. Ces variations temporelles semblent pouvoir trouver leur équivalent dans des différences entre pays fortement urbanisés et pays avec une population rurale ou agricole encore importante, comme la Roumanie, la Pologne ou la Grèce.

Une seconde tendance joue quant à elle dans le sens d'un renforcement de l'homogamie d'éducation : la massification scolaire. En effet, selon la théorie des systèmes éducatifs comme marché conjugués, l'allongement des études conduit à un renforcement de l'homogamie par le biais des rencontres dans le cadre scolaire. Nous avons vu au chapitre 6 que cet effet était réel dans le cas français, mais qu'il se révélait d'ampleur limitée ; en outre, massification et affaiblissement de l'homogamie d'éducation sont allés de pair en France depuis les années 1970. De ce fait, l'effet de la massification scolaire sur l'homogamie nous semble ambigu. Comme dans l'exemple français, il semble plausible que des processus entraînant une baisse de l'homogamie prennent place en même temps que la massification scolaire et en masquent les effets : une association au niveau individuel n'est pas nécessairement préservée lorsqu'elle s'applique à l'ensemble d'une société. En particulier, la massification scolaire entraîne une augmentation de l'entrecroisement des appartenances qui implique une diminution de l'homogamie.

Ces deux déterminants tenant aux lieux de rencontre sont mesurés par des variables structurelles simples calculées à partir de l'Enquête européenne sur les forces de travail. Le caractère agricole d'une région est appréhendé par la proportion de la population active exerçant dans le secteur primaire, défini comme les sections A et B de la Nomenclature statistique des activités économiques (NACE rev2), soit respectivement l'agriculture et la pêche, et l'extraction. Son caractère rural est mesuré au taux d'individus résidant dans une commune rurale, définies comme les unités administratives locales de niveau 2 qui ne se

3 Rappelons que nous recherchons les déterminants de l'homogamie relative : la taille du groupe des agriculteurs, fortement homogame, ne joue pas directement sur le niveau de l'homogamie d'une région.

rattachent pas à une grappe urbaine (qui sont des aires contiguës de carreaux de 1 km² dont la densité est supérieure à 300 habitants au km² et la taille totale supérieure à 5000 habitants). Enfin, la massification scolaire est mesurée par la proportion d'individus diplômés du supérieur, qui sont les plus susceptibles de rencontrer leur conjoint dans un cadre scolaire.

2.2 Entrecroisement des appartenances

L'entrecroisement des appartenances est supposé diminuer l'intensité de l'homogamie (Blau et Schwartz, 1984), d'abord en rendant difficile un choix homogame sur toutes les dimensions à la fois, et plus largement en atténuant les barrières entre groupes sociaux. Ces barrières peuvent tenir à des pratiques et des goûts différents, mais aussi à un cloisonnement des réseaux sociaux et des lieux de rencontre. Un indice d'un faible entrecroisement des appartenances est l'intensité du lien entre niveau d'éducation et profession. Celui-ci est mesuré ici à travers le logarithme de l'*odds ratio* absolu moyen (MAOR) calculé sur les données de l'Enquête sociale européenne à partir de tables croisant niveau d'éducation et groupe socioéconomique. Cet indicateur présente l'avantage de relier les deux dimensions de la stratification sociale retenues ici pour l'étude de l'homogamie, et dont l'importance est similaire pour le choix du conjoint : en cas de dissociation entre elles, les individus sont forcés de choisir l'un ou l'autre critère de choix du conjoint, diminuant l'homogamie de l'un, de l'autre ou des deux.

La mobilité sociale intergénérationnelle est un autre indicateur de l'entrecroisement des appartenances, dont on sait qu'il est négativement corrélé à l'intensité de l'homogamie d'éducation (Ultee et Luijkx, 1990 ; Katrňák, Fučík et Luijkx, 2012). On peut s'attendre à ce que cette variable présente un effet moins fort sur l'homogamie que la précédente, étant donné que, comme ces auteurs l'ont montré et comme nous l'avons observé dans le cas de la France, les caractéristiques des conjoints eux-mêmes importent plus que leurs origines sociales. Comme pour l'association entre niveau d'éducation et groupe socioéconomique, nous mesurons l'hérédité sociale en termes d'éducation (soit le contraire de la mobilité sociale) par le logarithme de l'*odds ratio* absolu moyen calculé à partir d'une table croisant niveau d'éducation des interrogés (hommes et femmes confondus) et de leurs pères à partir de l'Enquête sociale européenne. La même classification en trois catégories fondée sur la

Classification internationale type de l'éducation (CITE-97) est utilisée que dans l'Enquête européenne sur les forces de travail (voir chapitre précédent)⁴.

En outre, les proportions d'individus diplômés du supérieur, et de ceux appartenant aux classes supérieures (cadres et professions libérales et intellectuelles dans la classification ESeG) peuvent constituer une indication indirecte concernant l'expansion de ces groupes, et de la mobilité sociale structurelle qui en a découlé.

2.3 Niveau de vie

On peut penser que le niveau de vie joue un rôle dans le choix du conjoint. En effet, le risque associé au déclassement est moins important lorsque la situation matérielle des plus défavorisés leur assure un confort minimum ; au contraire, dans un contexte d'extrême pauvreté, une mésalliance peut s'avérer très coûteuse. Cette interprétation sous-tend l'hypothèse de l'amour romantique de J. Smits, W. Ultee et J. Lammers (1998) : l'amour gagne en importance dans une phase avancée du développement, comme un luxe que les individus peuvent se permettre lorsqu'un minimum de confort matériel est assuré.

Cette hypothèse peut être testée à l'aide d'au moins deux indicateurs. Premièrement, le niveau de développement économique, appréhendé par le produit intérieur brut (PIB) par habitant mesuré en parité de pouvoir d'achat, fournit une mesure globale du niveau de vie moyen dans la population considérée⁵. Cet indicateur est calculé au niveau régional par Eurostat. Nous retenons sa moyenne pour les années 2002 à 2010, et prenons pour référence la valeur 1 pour la moyenne de l'ensemble des régions sur la période.

Néanmoins, du fait des inégalités sociales, les ressources ne bénéficient pas nécessairement à l'ensemble de la population, et en particulier aux plus défavorisés. Pour cela, nous mesurons aussi quelle proportion de la population est privée du confort matériel

4 Le niveau d'éducation fourni par les données roumaines de la vague 4 de l'Enquête sociale européenne a dû être ignoré : en contradiction complète avec la vague 3 de la même enquête et avec les données de l'Enquête européenne sur les forces de travail 2011, il décrit une hérédité sociale et une homogamie parmi les plus faibles d'Europe. Les responsables de l'enquête, qui attribuent ce problème à l'absence d'utilisation de cartes-codes pour cette vague, ont ajouté un avertissement dans la documentation suite à notre signalement.

5 L'utilisation du revenu disponible net des ménages, qui tient compte des transferts sociaux et correspond donc plus directement au niveau de vie, fournit des résultats très proches. Seules les régions capitales présentent des différences significatives entre ces deux indicateurs, du fait de transferts importants vers les autres régions de leurs pays respectifs, mais sans pour autant modifier de manière sensible les résultats.

élémentaire. En effet, Eurostat calcule à partir de l'enquête Statistiques sur les revenus et les conditions de vie (SRCV/EU-SILC) le taux de personnes en situation de privation matérielle sévère. Cette situation recouvre des difficultés à payer son logement, à se chauffer, à faire face à des dépenses imprévues, à consommer de la viande ou une autre source de protéines régulièrement, à partir en vacances, ainsi que la non-possession d'équipements essentiels (télévision, machine à laver, voiture, téléphone). Cette variable n'est malheureusement pas définie à toutes les dates pour toutes les régions : elle n'est disponible qu'au niveau national pour l'Allemagne, la France, la Pologne, le Portugal et le Royaume-Uni ; l'Autriche et la Bulgarie ne fournissent que les taux nationaux avant 2008. Nous retenons la moyenne de cet indice pour les années 2005-2010.

2.4 Inégalités sociales et risque de mésalliance

Plus largement, le lien entre homogamie et inégalités fait l'objet de débats. En effet, celles-ci contribuent aux contrastes de styles de vie et de lieux fréquentés par les différents groupes sociaux. En outre, on peut penser que les réticences d'un individu à prendre pour conjoint une personne de statut social inférieur (et de son entourage à soutenir un tel choix) seront d'autant plus fortes que les écarts de revenus seront importants, et que les risques associés au déclassement seront forts. Ainsi, nous avons vu que de fortes inégalités de revenus selon le niveau d'éducation sont associées à une moindre fréquence des mises en couple entre groupes de diplôme (Torche, 2010 ; Monaghan, 2015). En généralisant ce résultat au niveau d'une population, on peut raisonnablement s'attendre à ce qu'un niveau élevé d'inégalités économiques soit associé à une plus forte homogamie⁶.

Comme nous l'avons discuté dans l'introduction de cette thèse, il est aussi possible qu'homogamie et inégalités soient liées par le biais de la tolérance plus ou moins grande d'une société par rapport à l'ampleur de ces dernières. D'un côté, une société dans laquelle les groupes sociaux se fréquentent peu pourrait faire preuve de plus de tolérance aux inégalités. De l'autre, un raisonnement opposé verrait dans la polarisation de la société – dont une forte homogamie est le signe – la condition du maintien d'une conscience de classe et donc d'une lutte contre les inégalités. C'est la logique de la spirale des classes sociales

6 Nous avons cependant vu au chapitre 3 que l'homogamie d'éducation ne renforçait pas directement les inégalités de salaires entre ménages. Si association il y a elle doit donc principalement passer par le renforcement des clivages entre groupes sociaux.

proposée par Louis Chauvel (2001) et déjà abordée au chapitre 2. Le lien entre homogamie et inégalités est donc *a priori* peu déterminé.

Plusieurs variables permettent de mesurer les inégalités économiques. L'indice de Gini mesure les inégalités relatives entre ménages, qui peuvent donner lieu à des différences de styles de vie. Nous retenons la valeur moyenne pour les années 2002-2010 des indices de Gini concernant le revenu des ménages après transferts sociaux tels que calculés par la Base de données standardisée des inégalités de revenus version 4 (*Standardized World Income Inequality Database*, Solt, à paraître), qui réunit et harmonise différentes sources dont notamment l'Étude luxembourgeoise sur les revenus (*Luxembourg Income Study*). Cet indicateur n'est malheureusement disponible qu'au niveau national.

D'autres indicateurs peuvent mesurer l'ampleur du risque de déclassement par la mise en couple. Tout d'abord, le taux de pauvreté relative. Défini comme la proportion de personnes vivant dans des ménages dont le revenu disponible après transferts sociaux se situe sous le seuil de pauvreté, défini à 60 % du revenu médian, il est calculé par Eurostat à partir des Statistiques sur les revenus et les conditions de vie (SRCV/EU-SILC). Nous retenons la valeur moyenne de cet indicateur pour les années 2005 à 2010⁷. Les données ne sont disponibles qu'au niveau national⁸.

2.5 Libéralisme des mœurs et contrôle social

Jusqu'ici, nos hypothèses ne différenciaient pas les préférences des individus et l'influence que peut exercer sur eux leur entourage. Cette distinction est dans de nombreux cas difficile à opérer du fait de l'intériorisation des normes sociales. On peut ainsi se demander si la liberté du choix du conjoint est synonyme d'une homogamie plus faible : l'un des apports essentiels de la sociologie à l'analyse de ce phénomène social est précisément d'avoir montré que l'amour n'est pas aveugle aux distinctions sociales. Néanmoins, il ne semble pas improbable que lorsque les individus peuvent choisir librement leur conjoint, ils auront moins tendance à faire primer le critère du statut social. Les affinités, si elles sont

7 À l'exception de la Roumanie (2007-2010).

8 Du fait de ce manque de précision au niveau géographique, nous avons eu recours au taux de chômage et au taux de chômage de longue durée, calculés par Eurostat au niveau régional, pour approcher la proportion de personnes dont la situation économique est très défavorable, ainsi que les difficultés à sortir de cette situation. Cependant, ces deux indicateurs n'apportant pas de véritable contribution à l'interprétation des résultats, nous ne les reprenons pas ici. Notons par ailleurs que le taux de privation matérielle, présenté ci-dessus, est calculé pour sa part au niveau régional.

corrélées aux appartenances sociales, le sont sans doute moins fortement que des stratégies conjugales fondées sur la seule considération du statut social ou de l'appartenance de groupe. C'est ce que nous enseignent les multiples cas de conflit entre parents et enfants dans les sociétés où le choix du conjoint est soumis à l'approbation des premiers (Flandrin, 1975 ; Macfarlane, 1986). Cette hypothèse nous semble ainsi devoir être testée. Pour cela, deux indicateurs de la prégnance des normes sociales sur les individus peuvent être calculés.

Premièrement, le contrôle social peut passer par des contacts fréquents et significatifs avec la famille – et notamment avec les parents – des jeunes gens en âge de se mettre en couple. Par exemple, l'influence parentale pourrait être plus forte en Europe du Sud, où le départ du domicile parental est tardif et a fréquemment lieu après une période de vie en couple non cohabitante au cours de laquelle les conjoints sont amenés à fréquenter leurs belles-familles (Reher, 1998 ; Bernardi et Oppo, 2011) ; et plus forte encore en Europe Centrale et Orientale, où la cohabitation du jeune couple chez les parents est commune (Billari, Philipov et Baizán, 2001). Nous ne disposons malheureusement pas d'informations concernant ces contacts, ni de renseignements indirects portant sur l'âge de décohabitation ou la proximité géographique de l'entourage de l'interrogé au moment où celui-ci s'est mis en couple.

À défaut de pouvoir appréhender précisément les liens entre l'individu et son entourage, la proportion de ménages comprenant plusieurs couples (ménages multiples dans la typologie de Peter Laslett) peut être prise comme le signe d'une cohabitation fréquente des générations au-delà de l'âge de mise en couple de la jeune génération⁹. Cet indicateur est naturellement incapable de saisir une partie importante de liens familiaux, qui dépassent fréquemment les limites du ménage (Bonvalet et Lelièvre, 1995). Néanmoins, il permet d'identifier le cas le plus extrême d'influence parentale sur le choix du conjoint : la cohabitation des générations suppose un contact fréquent et prolongé entre l'un des conjoints et ses beaux-parents avant la mise en ménage, mais aussi dans la perspective de relations durables. Ceci semble impliquer une plus grande capacité des parents à influencer

9 Des informations concernant l'âge de départ du domicile parental sont disponibles dans la vague 3 de l'Enquête sociale européenne (Billari et Liefbroer, 2010), qui ne couvre malheureusement pas l'ensemble des pays étudiés. Notons par ailleurs que notre utilisation de la proportion de ménages multiples diffère de celle retenue par Emmanuel Todd (1990, p. 37), qui vise, en croisant degré d'urbanisation et proportion de ménages complexes, à identifier les traces d'un système familial autoritaire. Notre approche consiste plus simplement à rechercher un effet potentiel de contacts concrets entre générations, la question des valeurs étant traitée ci-dessous à partir de questions directes.

les choix de leurs enfants (Goode, [1964] 1982, chap. 11). Le taux de ménages multiples est calculé à partir de l'Enquête européenne sur les forces de travail, selon la définition du couple présentée plus haut.

Deuxièmement, les valeurs plus ou moins individualistes d'une société peuvent refléter une plus grande propension à choisir son conjoint indépendamment des influences de ses proches, mais aussi, vu du côté parental, à laisser ses enfants choisir librement leur conjoint. Encore une fois, il paraît donc délicat de séparer complètement préférences des individus et contrôle de leur entourage. L'Enquête sociale européenne contient une série de questions visant à mesurer le positionnement des interrogés sur les dix valeurs fondamentales de la théorie de Shalom Schwartz (Davidov, Schmidt et Schwartz, 2008). Parmi ces dix valeurs, dont certaines sont susceptibles d'entrer en conflit alors que d'autres sont plus congruentes entre elles, plusieurs sont adaptées à l'appréhension de l'importance des normes sociales pour le choix du conjoint. Chacune donne lieu à une question dans laquelle l'interrogé se voit décrire un individu, dont il doit déclarer s'il se sent plus ou moins proche sur une échelle de 1 à 6. La stimulation (*stimulation*) mesure l'importance accordée à la nouveauté et aux défis ; la question correspondante est formulée ainsi dans le questionnaire français : « *Il recherche l'aventure et aime prendre des risques. Il désire une vie excitante.* » L'autonomie (*self-direction*) correspond à la propension à penser et à agir librement et par soi-même : « *C'est important pour lui de décider tout seul de ce qu'il va faire. Il aime être libre et ne pas dépendre des autres.* » Au contraire, la tradition (*tradition*) désigne le respect des coutumes, dont notamment celles de la famille : « *Les traditions sont importantes pour lui. Il essaie de suivre les coutumes transmises par la religion et la famille.* » Le conformisme (*conformity*) représente l'importance de suivre les règles : « *Il pense que les gens devraient faire ce qu'on leur dit. Il pense que les gens devraient toujours suivre les règles établies, même si personne ne les surveille.* » Nous retenons pour chaque question la moyenne des notes des individus pour toutes les années d'enquête.

À ces valeurs générales, il nous a paru intéressant d'adjoindre un indicateur du libéralisme culturel qui concerne une attitude plus concrète et plus directement liée au domaine familial. L'Enquête sociale européenne contient dans toutes ses vagues une question portant sur un indicateur classique de ce point de vue (Bréchon et Galland, 2010,

p. 17) : la tolérance à l'homosexualité¹⁰. Cette question est traduite dans l'enquête française de la manière suivante : « *Les homosexuels hommes et femmes devraient être libres de vivre leur vie comme ils le souhaitent* ». Les enquêtés peuvent répondre sur une échelle de 1 à 5 ; comme précédemment, nous retenons la moyenne des notes des individus pour toutes les années d'enquête.

Enfin, pour compléter ces informations reposant sur des déclarations, nous avons inclus un indicateur reposant sur un comportement démographique : la proportion de couples mariés. Cette proportion est considérée comme un indice du degré d'institutionnalisation de la famille. Nous avons vu au chapitre 1 que plusieurs études avaient mis en évidence une moindre homogamie parmi les couples cohabitants que parmi les couples mariés. En passant du niveau individuel au niveau d'une société, nous pouvons donc raisonnablement nous attendre à ce que les régions dans lesquelles la cohabitation est la plus développée soient celles dans lesquelles l'homogamie est la plus faible. La proportion de couples mariés est calculée à partir de l'Enquête européenne sur les forces de travail.

2.6 Religion

Comme on l'a vu au chapitre précédent, la religion dominante d'une société a été utilisée par plusieurs des travaux existants comme indication indirecte des valeurs prévalant dans une société, et notamment de l'importance accordée au respect de la tradition. Grâce aux données de l'Enquête sociale européenne, nous pouvons appréhender directement les valeurs des individus. De ce fait, la religion ne nous semble pas devoir jouer un rôle théorique propre, au-delà des valeurs elles-mêmes. On sait par exemple qu'aussi bien à l'échelle individuelle qu'à celle d'un pays, l'athéisme et le protestantisme sont associés à un niveau d'individualisation des valeurs plus élevé que le catholicisme et surtout que le christianisme orthodoxe (Bréchon, 2013).

Néanmoins, nous incluons des variables concernant le degré de religiosité, de manière à confirmer notre affirmation, et à éventuellement résumer les oppositions entre systèmes de valeurs par une caractérisation religieuse. Nous calculons donc à partir de l'Enquête sociale européenne le degré de religiosité moyen des individus concernant quatre religions :

10 Nous tenons à remercier Yannick Lemel pour nous avoir signalé ce point. D'autres questions portant sur la famille ont été posées dans certaines vagues de l'enquête, mais elles ne couvrent pas toutes les régions étudiées.

catholique, protestante, orthodoxe et autre (minorités religieuses). Ce degré est déclaré par les enquêtés sur une échelle de 0 à 10 ; les individus n'ayant pas déclaré appartenir à la religion considérée se voient attribuer un degré de religiosité nul pour cette religion. Notons qu'une manière différente de mesurer la religiosité, en retenant la fréquence de la pratique, fournit des résultats très proches (corrélation supérieure à 0,9 pour toutes les religions).

3 Le statut des femmes et les déterminants de l'hypergamie féminine

3.1 Activité féminine et choix du conjoint : un lien débattu

Le lien entre évolution du statut des femmes et choix du conjoint est débattu : la question est notamment de savoir si l'augmentation de l'activité professionnelle des femmes doit se traduire par un transfert de l'hypergamie féminine vers l'homogamie. Il est donc nécessaire de discuter simultanément de ces deux aspects du choix du conjoint.

Comme nous l'avons vu au chapitre 4, la moindre dépendance économique des femmes leur permet d'accorder moins d'importance aux perspectives de carrière de leur conjoint, entraînant donc une baisse de l'hypergamie. Mais les hommes pourraient tenir plus compte qu'avant de ces aspects (Oppenheimer, 1988 ; Bozon, 1990 ; Buss *et al.*, 2001), ce qui entraînerait un renforcement de l'homogamie (Mare, 1991 ; Blossfeld et Timm, 2003). Nous avons expliqué en introduction pourquoi cette dernière hypothèse nous semblait cependant partielle. En retenant une lecture stratégique, elle néglige en effet l'importance du statut social et du capital culturel au-delà de la seule question économique – sans parler même du capital économique hérité par les femmes. Si l'on se place toujours dans un modèle de choix stratégique du conjoint, même dans un modèle fortement genré dans lequel les femmes n'auraient aucune activité économique à l'extérieur du ménage, la représentation sociale du ménage et l'éducation des enfants – par lesquelles passe la reproduction sociale – sont très largement tributaires du statut social de la femme, ainsi que nous l'avons développé au chapitre 4. Même sans se placer dans ce cadre théorique, la simple similarité des fréquentations et des styles de vie suffit à entraîner une forte homogamie.

Par conséquent, l'augmentation de l'activité des femmes n'a pas nécessairement pour effet de renforcer l'homogamie, qui peut aussi être très élevée dans un autre contexte du fait

de l'action d'autres facteurs. Au contraire, on peut penser que la plus grande indépendance des femmes que leur offre leur activité professionnelle leur permet d'accorder moins d'importance aux perspectives de carrière des conjoints potentiels, et plus aux aspects relationnels et esthétiques (Bozon, 1990 ; Buss *et al.*, 2001). De fait, abondant dans le sens des résultats de Jeroen Smits et Hyunjoon Park (2009) ou de David Monaghan (2015) présentés au chapitre précédent, notre étude a montré que les évolutions temporelles de ces deux phénomènes en France constituaient un contre-exemple flagrant à la théorie postulant une association positive entre activité féminine et homogamie. Il s'agit maintenant de confirmer ces résultats dans le cas européen.

Nous retenons comme indicateur du statut professionnel des femmes la proportion d'actives (occupées ou chômeuses) parmi les femmes âgées de 30 à 59 ans, calculée à partir de l'Enquête européenne sur les forces de travail. La fenêtre d'âge relativement restreinte permet de nous prémunir contre une interférence des variations entre régions de l'âge de fin d'études d'un côté, et de l'âge de départ en retraite de l'autre.

3.2 Statut des femmes et hypergamie

Par ailleurs, si l'hypergamie féminine relative est bien le reflet de la domination masculine, on peut s'attendre à ce que les différents indicateurs de la condition féminine lui soient associés. Il s'agit tout d'abord de la position des femmes dans le système éducatif et sur le marché du travail. L'inégalité entre hommes et femmes du point de vue de l'éducation est mesurée comme la probabilité qu'en choisissant un homme et une femme de niveaux d'éducation différents au hasard dans la population, la femme soit plus diplômée que l'homme (Esteve, García-Román et Permanyer, 2012, p. 540)¹¹. Cet indicateur est égal à 0 si tous les hommes sont plus diplômés que la femme la plus diplômée (inégalité totalement en faveur des hommes), et à 1 si l'inverse est vrai (inégalité totalement en faveur des femmes). Il prend la valeur 0,5 quand les deux sexes se distribuent de manière identique par niveau d'éducation. Cet indicateur est calculé à partir de l'Enquête européenne sur les forces de travail, en utilisant la classification en trois catégories du chapitre précédent.

11 Cet indice correspond en fait à la différence entre la proportion de couples à hypogamie féminine et celle de couples à hypergamie féminine qui seraient observées parmi les couples hétérogames si les couples se formaient totalement au hasard et si le célibat était nul (ou plus précisément, à la moitié de 1 plus cette différence). Il est donc très proche de la mesure de l'hypergamie absolue utilisée au chapitre précédent.

Cette démarche est reprise pour la mesure des inégalités entre hommes et femmes sur le marché du travail, toujours à partir de l'Enquête européenne sur les forces de travail. En plus du taux d'activité féminine, déjà présenté plus haut comme déterminant de l'homogamie, nous calculons la probabilité qu'en choisissant un homme et une femme au hasard dans la population âgée de 30 à 59 ans, et si les groupes d'appartenance de deux individus sont ordonnés hiérarchiquement, la femme ait un statut supérieur à l'homme. Les groupes sont classés en utilisant la matrice d'hypergamie définie au chapitre précédent (Tableau 7.2, p. 416), avec cependant pour différence notable que les inactifs sont considérés comme ayant un statut inférieur à tous les autres groupes, de manière à tenir compte des différences de taux d'activité féminine entre régions.

On peut s'attendre à ce que ces indicateurs directs du statut des femmes dans une société soient liés à de nombreuses autres dimensions déjà évoquées à propos de l'homogamie. Ainsi, une plus grande importance accordée à la tradition pourrait être associée à une conception fortement différenciée des rôles masculins et féminins. Malheureusement, des indicateurs plus directement liés à la question des rôles de genre (comme la priorité qu'une femme doit accorder à la famille plutôt qu'à l'activité professionnelle, ou l'acceptabilité de la contraception et du fait de ne pas avoir d'enfants), présents dans certaines vagues de l'Enquête sociale européenne, ne sont pas disponibles pour tous les pays étudiés. Les analyses sur un ensemble restreint de régions montrent cependant que ces questions portant sur les attitudes apportent peu à un modèle qui contient déjà des informations concernant à la fois des valeurs générales et des pratiques concrètes (libéralisme culturel, proportion de couples mariés, tolérance à l'homosexualité).

Par ailleurs, on peut penser que la proportion de couples mariés figure parmi les déterminants de l'hypergamie féminine. En effet, comme nous l'avons défendu en introduction, le mariage a pu être vu par les femmes qui rejetaient la forte division des rôles au sein du couple comme le symbole de ce modèle familial. Il constitue par ailleurs le cadre privilégié de la spécialisation des rôles, favorisé par un engagement à long terme dans le couple et des garanties juridiques entourant le divorce ou le décès de l'un des conjoints.

4 Méthode : la régression PLS

Afin de tester les hypothèses que nous venons de développer, nous recourons à une méthode originale qui semble particulièrement adaptée à notre problématique : la régression PLS (Tenenhaus, 1998 ; Abdi, 2010), pour régression des moindres carrés partiels (*Partial Least Squares*) ou régression par projection sur structures latentes (*Projection to Latent Structures*). Cette technique est à mi-chemin entre analyse en composantes principales (ACP) et régression linéaire multiple ; son interprétation se rapproche pour cela de celle des analyses factorielles.

4.1 Principe : une analyse factorielle tournée vers l'explication

La régression PLS étend le principe de l'analyse factorielle vers la démarche propre à la régression en cherchant à expliquer un ensemble de variables (une ou plusieurs) par un autre ensemble de variable indépendantes. Alors que l'analyse factorielle consiste à identifier les composantes (ou dimensions, ou encore axes) permettant de résumer le mieux l'association au sein d'un ensemble de variables, la régression PLS cherche les composantes (ou variables latentes) maximisant la description à la fois des variables explicatives, des variables expliquées, et de la corrélation entre ces deux ensembles de variables¹². Les variables explicatives se trouvent résumées par une série de composantes non corrélées entre elles, sur lesquelles chaque région se voit attribuer un score, et qui sont à leur tour introduites dans une régression cherchant à modéliser les variables expliquées (Encadré 8.1). L'importance de chaque composante peut être évaluée à travers la proportion de la variance de chacune des variables expliquées, mais aussi à celle de l'ensemble des variables explicatives, dont elle rend compte.

Une régression PLS retenant l'ensemble des composantes décrit parfaitement l'ensemble des variables expliquées, et se ramène donc à une régression linéaire multiple classique. Comme en analyse factorielle, l'intérêt de la méthode est de réduire la complexité des données en s'en tenant au nombre de composantes nécessaires à une modélisation correcte

12 La régression PLS se distingue ainsi de la combinaison de l'ACP et de la régression linéaire multiple connue sous le nom de régression sur composantes principales, puisqu'elle identifie les composantes qui sont le mieux à même de décrire la variable expliquée, alors que la seconde méthode choisit les composantes de manière à décrire uniquement les variables explicatives. Elle diffère pour les mêmes raisons de la projection en variables supplémentaires des variables expliquées sur une analyse factorielle fondée sur les variables explicatives.

des variables expliquées. Quel que soit le nombre de composantes retenues, la prédiction des variables expliquées sera optimale pour ce niveau de complexité. Le choix du nombre de composantes se fonde généralement sur l'évaluation de la qualité de prédiction du modèle dans le cadre d'une validation croisée. Cette technique vise à éviter d'accorder trop d'importance à des particularités tenant à seulement quelques régions, en estimant successivement le modèle après avoir exclu une partie des régions. Les composantes qui ne permettent pas d'améliorer la prédiction du modèle dans cette situation ne sont pas retenues.

Encadré 8.1 : L'algorithme de la régression PLS

L'estimation d'une régression PLS est réalisée en appliquant de manière répétée une série d'étapes permettant d'identifier successivement les différentes dimensions. Cette procédure est d'abord appliquée aux données originales pour identifier la première dimension, puis réitérée sur le résidu des données non expliqué par les dimensions déjà identifiées.

La première suite d'étapes consiste à estimer deux séries de scores latents t_1 et u_1 , correspondant à la première dimension, qui sont respectivement les combinaisons linéaires des variables explicatives x et des variables expliquées y , de manière à ce que leur covariance soit maximale. Ces deux séries de scores (un pour chaque observation – ici, région) s'écrivent :

$$t_1 = \sum_{j=1}^J w_{1j} x_j \text{ et } u_1 = \sum_{k=1}^K c_{1k} y_k$$

Dans cette équation, x_j est la variable explicative numéro j et y_k la variable expliquée numéro k ; w_{1j} et c_{1k} sont les poids mesurant la contribution de chaque variable à la construction de la première dimension. Nous ne détaillons pas les étapes permettant d'estimer ces scores et ces poids.

Une fois les deux séries de scores latents identifiées, on soustrait aux variables d'origine la partie expliquée par la première composante (déflation). La soustraction de cette composante aux variables originales est réalisée en régressant les variables correspondantes sur les scores latents t :

$$x_j = t_1 p_{1j} + x_{2j} \text{ et } y_k = t_1 c_{1k} + y_{2k}$$

Cette régression fournit les coefficients p , qui permettent de passer des scores latents t aux variables explicatives x . Pour leur part, du fait de l'asymétrie entre variables explicatives et expliquées, les coefficients c correspondent à ceux déjà calculés à l'équation précédente ; on remarque que les scores latents u jouent un rôle très limité.

Les résidus de ces deux régressions x_{2j} et y_{2k} , qui représentent la part des variables dont le modèle ne rend pas encore compte, prennent la place des variables originales x_j et y_k . La procédure est alors réitérée pour l'identification de la deuxième dimension, l'objectif étant toujours d'identifier deux séries de scores latents dont la covariance est maximale :

$$t_2 = \sum_{j=1}^J w_{2j} x_{2j} \text{ et } u_2 = \sum_{k=1}^K c_{2k} y_{2k}$$

Notons que l'algorithme garantit que la corrélation entre les scores latents t correspondant à deux dimensions différentes est nulle :

$$\text{cor}(t_h, t_{h'}) = 0 \text{ pour tout } h \neq h'$$

Une fois toutes les dimensions estimées, une étape finale est nécessaire pour calculer le poids de chaque variable originale dans la construction des dimensions successives. En effet, les poids w ci-dessus mesurent la contribution des *résidus* des variables calculés à chaque étape à la dimension correspondante. Afin de permettre l'interprétation directement à partir des données originales, on calcule donc une nouvelle série de poids w^* . On cherche en effet à se ramener à l'équation valable pour la première dimension, pour laquelle on a utilisé les variables originales et non les résidus, pour toute dimension quelconque h :

$$t_h = \sum_{j=1}^J w_{hj}^* x_j$$

Les poids w^* sont donc les coefficients de la régression des scores latents t sur les variables explicatives originales x . Du fait de la corrélation nulle entre scores latents t , les poids c sont quant à eux à la fois les coefficients de corrélation entre les variables

expliquées y et les scores latents t , et les coefficients de la régression de ces variables sur ces mêmes scores :

$$\text{cor}(y_k, t_h) = c_{hk} \text{ et } y_k = \sum_{h=1}^H t_h c_{hk}$$

Les coefficients de régression b liant chaque variable explicative x_j à chaque variable expliquée y_k (équivalents aux coefficients d'une régression linéaire classique) sont calculés à partir des poids w^* et c . Ils sont simplement égaux à la somme du produit de ces deux séries de poids effectué pour chacune des H dimensions du modèle :

$$b_{jk} = \sum_{h=1}^H w_{hj}^* c_{hk}$$

Les poids des variables présentés dans les graphiques plus bas correspondent aux coefficients w^* et c . Leur interprétation est détaillée dans le corps du texte.

4.2 Du « problème du petit N » à une analyse « avec noms propres »

L'intérêt de la régression PLS pour notre étude est double. Premièrement, comme dans la plupart des études comparatives au niveau macro, notre échantillon de régions est de taille limitée par rapport au nombre de variables qu'il semble intéressant d'inclure dans le modèle. Cette difficulté est généralement désignée sous le nom de « problème du petit N » (N désignant la taille de l'échantillon, cf. Lijphart, 1971 ; Goldthorpe, 1997). Même si le nombre de régions reste supérieur au nombre de variables introduites, la forte corrélation qui existe entre ces dernières (par exemple, entre niveau de développement, libéralisme culturel et urbanisation), et plus encore entre leurs combinaisons (multicolinéarité) génère une instabilité des coefficients. Cette difficulté n'est pas seulement technique, elle est plus fondamentalement logique et théorique. Lorsque les caractéristiques des régions étudiées sont trop liées les unes aux autres, il n'apparaît pas scientifiquement valide de penser pouvoir distinguer leurs effets : l'approche « toutes choses égales par ailleurs » n'a pas grand sens lorsque l'échantillon ne contient aucun cas dans lesquels deux variables explicatives divergent de manière notable.

Nous ne renonçons cependant pas à chercher à expliquer les variations de l'homogamie et de l'hypergamie. Simplement, au lieu de ne retenir *a priori* qu'un nombre réduit de

variables considérées comme déterminantes – au risque de réaliser un choix arbitraire, de déformer le modèle théorique et de biaiser les estimations –, nous incluons toutes les variables pertinentes disponibles, la régression PLS permettant d'identifier les méta-facteurs qui, synthétisant l'ensemble des variables considérées, sont les meilleurs prédicteurs de l'homogamie et de l'hypergamie. Seuls ces méta-facteurs fondamentaux sont contrôlés dans l'analyse de régression : comme en analyse factorielle, l'interprétation consiste alors à étudier l'association brute entre ces dimensions et chacune des variables.

L'impossibilité de distinguer les effets de trop nombreux déterminants à partir d'une réalité à la diversité limitée est ainsi pleinement reconnue. Deux explications concurrentes du phénomène modélisé seront considérées comme également valides si elles sont trop proches sur tous les méta-facteurs pour être séparées à partir des données disponibles. Cette démarche nous semble préférable à celle qui consisterait à choisir de tester une hypothèse plutôt qu'une autre pour la seule raison que les deux ne peuvent être traitées en même temps dans un modèle donné.

L'approche au fondement des techniques d'analyse factorielle, consistant à synthétiser les multiples déterminants en un nombre réduit de dimensions, présente en outre l'avantage significatif de permettre la projection simultanée des variables *et* des régions sur ces axes. La régression PLS permet ainsi à la fois d'abstraire les régions européennes de leurs particularités en les rendant comparables par une mise en variables, et de préserver leurs « noms propres » de manière à ne jamais perdre de vue les contextes concrets dans lesquels les variables s'actualisent (Shalev, 2007, p. 264 ; Esping-Andersen, 2007). Cette méthode représente donc une réponse possible au dilemme classique des études comparatives entre prétention à la généralisation théorique et perte de sens historique liée à l'invisibilisation des cas sur lesquels l'analyse se fonde (Goldthorpe, 1997), et plus largement à l'opposition entre « langage des groupes » et « langage des variables » (Desrosières, 1995, 2001).

4.3 La modélisation de plusieurs variables

Le deuxième intérêt majeur de la régression PLS est de rendre possible la modélisation simultanée de plusieurs variables expliquées. Notre étude porte justement sur deux dimensions différentes de l'homogamie et de l'hypergamie – le diplôme et le groupe socioéconomique – dont nous avons toutes les raisons de penser qu'elles ont des

déterminants communs. Plutôt que de procéder à quatre modélisations séparées, la régression PLS nous permet de repérer simultanément les points communs et les différences de ces deux dimensions, ce qui permet de répondre assez directement à nos interrogations.

4.4 Estimation

Le modèle est estimé dans l'environnement R (R Core Team, 2014) à l'aide du paquet pls (Mevik, Wehrens et Liland, 2013). Les pondérations n'ont pas pu être appliquées, n'étant pas prises en charge pour un modèle à plusieurs variables expliquées. Variables expliquées – homogamie et hypergamie – comme variables explicatives sont standardisées (centrées et réduites), afin de les rendre commensurables bien qu'elles ne soient pas exprimées dans la même unité et de ne pas attribuer plus d'importance aux dimensions du choix du conjoint présentant les plus grandes variations entre pays. Ainsi, l'ampleur des effets est comparable d'une variable à l'autre, une unité supplémentaire correspondant à un écart-type entre régions de l'échantillon. Il n'est cependant pas inutile de garder en mémoire que les variations de l'homogamie et de l'hypergamie sont plus fortes du point de vue de l'éducation que de celui du groupe socioéconomique.

5 Résultats : des déterminants largement communs

L'étude du pouvoir explicatif des différentes dimensions, à travers la proportion de variance dont elles rendent compte (R^2 , voir Tableau 8.2), fait apparaître quelques résultats saillants. Premièrement, les variations de l'homogamie relative d'éducation s'expliquent à 69 % par une seule dimension dont nous verrons ci-dessous qu'elle correspond au niveau de développement, mais surtout au degré d'ouverture sociale et de libéralisme culturel.

Deuxièmement, les variables que nous avons retenues expliquent nettement mieux les variations de l'homogamie que celles de l'hypergamie. Ainsi, si nous retenons, la validation croisée nous y incite, les quatre premières dimensions, le modèle explique respectivement 26 % et 31 % des variations de l'hypergamie d'éducation et socioéconomique, mais 78 % de l'homogamie socioéconomique et même 89 % de l'homogamie d'éducation. Les dimensions au-delà de la quatrième expliquent chacune une part légèrement décroissante des variations de l'hypergamie, sans qu'un palier très net se détache ; elles ne permettent pas d'améliorer

la prédiction du modèle dans le cadre d'une validation croisée. Cette difficulté à modéliser l'hypergamie relative tient probablement en grande partie au fait que sa mesure est entachée d'approximations importantes, comme on l'a vu au chapitre précédent à l'ampleur des intervalles de confiance. Si cette imperfection de mesure, en introduisant des erreurs aléatoires, réduit le pouvoir explicatif des variables à supposer qu'elles aient véritablement un effet, elle n'introduit normalement pas de biais dans l'identification de cet effet.

Notons finalement que les dimensions retenues décrivent 72 % de la variance des variables explicatives, le premier axe étant de loin, comme pour l'homogamie d'éducation, le plus important.

Tableau 8.2 : Proportion de variance expliquée par les dimensions (R² en %)

	1	2	3	4	5	Total 4 dim.
Homogamie d'éducation	68,9	16,1	2,4	0,0	1,5	88,9
Homogamie socioéconomique	40,7	0,7	26,8	9,5	0,2	78,0
Hypergamie d'éducation	7,3	3,0	2,8	3,9	9,4	26,3
Hypergamie socioéconomique	7,8	16,1	0,9	2,7	3,7	31,1
Variables explicatives	39,0	10,0	8,5	9,0	5,6	72,0

Nous présentons successivement les quatre dimensions retenues. Les dimensions d'une régression PLS s'interprètent de manière similaire à une analyse factorielle, mais avec quelques particularités (Tenenhaus, 1998, p. 87). Ici, les scores (cordonnées) attribués aux régions sur chaque dimension représentent des variables latentes qui font le lien entre variables explicatives et variables expliquées (homogamie et hypergamie). Rappelons que ces scores ont été choisis de manière à décrire au mieux, avec une importance décroissante, ces deux ensembles et la relation qui les unit. La régression PLS correspond à deux étapes séparées : la première permet de passer des variables explicatives aux scores latents attribués aux régions ; la seconde, de ces scores latents à la prédiction proprement dite des variables expliquées. L'interprétation des dimensions de la régression PLS est donc réalisée à partir des coefficients correspondant à ces deux étapes. Les variables étant standardisées, les poids des variables explicatives mesurent l'importance prise par chaque variable dans la

création de la dimension considérée, alors que les poids des variables expliquées mesurent l'importance de chaque dimension dans leur modélisation.

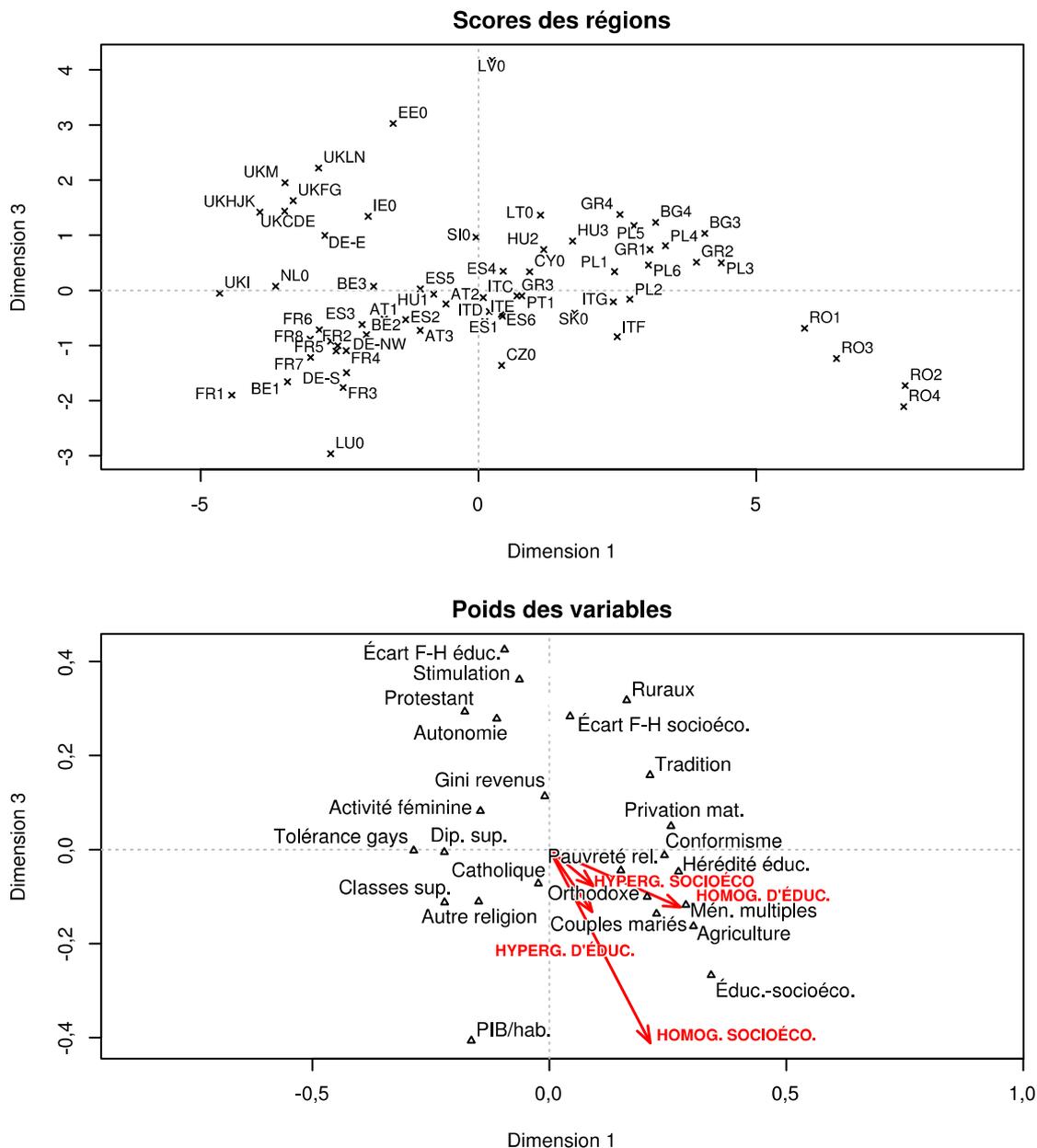
Le lien entre une variable explicative et une variable expliquée se mesure à la somme des produits des coordonnées respectives de ces variables sur chaque axe (Encadré 8.1, p. 476). Visuellement, pour une paire de dimensions représentée sur un plan, cette somme est égale à la distance à l'origine de la projection orthogonale du point correspondant à la variable sur la flèche correspondant à la variable expliquée. Ainsi, une variable explicative fortement corrélée à une variable expliquée aura un angle proche de celle-ci, et sera très éloignée de l'origine ; une variable explicative négativement corrélée à une variable expliquée sera elle aussi éloignée de l'origine, mais formera un angle proche de 180° avec celle-ci. Une variable explicative faiblement corrélée avec une variable expliquée formera un angle droit avec celle-ci, ou sera proche de l'origine ; dans ce dernier cas, elle sera en outre peu associée à l'ensemble des variables explicatives sur le plan considéré. Le lien total entre une variable expliquée et une variable explicative est égal à la somme des liens correspondant à chaque paire de dimensions – soit ici aux deux plans présentés plus bas.

5.1 Les déterminants communs de l'homogamie et de l'hypergamie : libéralisme culturel, ouverture sociale et niveau de développement

Nous commençons par présenter les dimensions 1 et 3 (Figure 8.1), qui représentent les déterminants largement communs à l'homogamie et à l'hypergamie aussi bien en termes éducatifs que socioéconomiques, quoique avec des nuances intéressantes¹³. La première dimension (horizontale) est par construction celle qui permet le mieux de décrire les variations entre régions des variables explicatives, des quatre phénomènes modélisés, et du lien entre les deux. Conformément aux théories de la modernisation, elle est bien liée au niveau de développement ; mais ce n'est pas sa définition essentielle : ce sont avant tout l'ouverture de la structure sociale et le libéralisme culturel qui contribuent à sa construction. Toutes les composantes de l'homogamie et de l'hypergamie sont négativement corrélées à ces deux caractéristiques (à droite de l'axe), mais c'est particulièrement le cas de l'homogamie d'éducation – dont ce premier axe explique près de 70 % des variations.

13 La grande importance prise par les facteurs communs à l'homogamie et à l'hypergamie n'est pas un artefact lié à leur modélisation conjointe : on retrouve la même structure du plan en modélisant les deux phénomènes séparément.

Figure 8.1 : Les dimensions 1 et 3 de la régression PLS



Source : Enquête européenne sur les forces de travail 2011 ; Enquête sociale européenne, vagues 1 à 5 (2002-2010) ; Eurostat ; Standardized World Income Inequality Database.

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour le groupe socioéconomique) pour la mesure de l'homogamie et de l'hypergamie ; ensemble de la population de plus de 15 ans pour les autres variables (sauf exception, voir le corps du texte).

Cf. Tableau 7.5 (p. 454) pour la correspondance entre codes et noms des régions.

La variable contribuant le plus à la première dimension confirme l'importance de l'entrecroisement des appartenances dans la détermination du niveau de l'homogamie : il s'agit en effet de l'intensité du lien entre niveau d'éducation et groupe socioéconomique (noté *Éduc.-socioéco.*), qui est associé à plus d'homogamie et d'hypergamie (côté droit de l'axe). Un autre indicateur d'une faible ouverture sociale, l'intensité du lien entre le niveau d'éducation des parents et celui de leurs enfants (*Hérédité éduc.*), est lui aussi parmi les plus contributeurs au côté droit de l'axe. Enfin, un troisième indicateur que la théorie relie aux deux précédents se situe au même niveau sur la première dimension : la proportion d'actifs exerçant dans l'agriculture (*Agriculture*).

Deux indicateurs d'importance similaire, qui relèvent quant à eux du domaine du libéralisme culturel et du contrôle social, viennent juste après : tout d'abord, toujours du côté droit (homogamie et hypergamie élevées), la proportion de ménages multiples, indice de l'importance de l'influence familiale sur le choix du conjoint ; ensuite, du côté gauche de l'axe, apparaît la tolérance à l'homosexualité (*Tolérance gays*), signe de libéralisme des mœurs. À l'inverse, d'autres indicateurs d'un faible libéralisme culturel, du point de vue des valeurs autant que des comportements, apparaissent déterminants du côté droit de l'axe : l'importance accordée au respect des règles (*Conformisme*) et de la tradition (*Tradition*), ainsi que la proportion de couples mariés (*Couples mariés*).

Les facteurs économiques ne viennent qu'ensuite, à travers un indicateur croisant niveau de vie et inégalités : la proportion de personnes en situation de privation matérielle sévère (*Privation mat*), associée au côté droit de l'axe. Le niveau de développement proprement dit (*PIB/hab.*), associé au côté gauche de l'axe, apparaît moins déterminant pour le choix du conjoint.

Soulignons pour finir la contribution très similaire de la proportion de diplômés du supérieur (*Dip. sup.*) et de classes supérieures (*Classes sup.*) du côté gauche de l'axe, qui semblent confirmer le lien négatif entre entrecroisement des appartenances découlant de l'expansion numérique de ces groupes et homogamie. Comme nous l'avions défendu, la massification scolaire ne s'accompagne pas d'un renforcement de l'homogamie d'éducation, l'effet de l'entrecroisement des appartenances l'emportant sur l'augmentation de la fréquence des rencontres dans le cadre scolaire.

Le degré de religiosité croisé avec la religion dominante apporte un éclairage complémentaire à cette première dimension. Le protestantisme (*Protestant*) et les autres religions (*Autre religion*) sont associés à moins d'homogamie et d'hypergamie (côté gauche de l'axe), à l'inverse du christianisme orthodoxe (*Orthodoxe*) ; le catholicisme (*Catholique*) apparaît relativement peu marqué sur cet axe.

Ainsi, la stratification sociale et les déterminants culturels jouent un rôle plus net dans la détermination de l'homogamie et de l'hypergamie que le niveau de développement économique en lui-même. Ce résultat permet de préciser les mécanismes par lesquels la modernisation est liée au choix du conjoint : c'est avant tout du fait d'un entrecroisement des appartenances plus fort, d'un moindre contrôle social et d'un moindre attachement à la tradition (à la fois sur le plan des valeurs et des pratiques) que le développement entraîne une diminution de l'homogamie et de l'hypergamie.

Le troisième axe (vertical) permet de tenir compte de certains cas particuliers significatifs. En effet, on remarque que la plupart des régions se distribuent sur un axe oblique allant de l'angle inférieur gauche à l'angle supérieur droit : pour ces régions, le troisième axe n'apporte guère d'informations par rapport au premier. En revanche, ce nouvel axe permet de tenir compte de la spécificité de deux groupes de régions déjà identifiés en partie au chapitre précédent. En bas à droite, les régions roumaines se distinguent par une très forte homogamie et une très forte hypergamie féminine ; en haut à gauche, des régions du Nord de l'Europe (Allemagne de l'Est, Estonie, Irlande, Lettonie et Royaume-Uni ; la Lituanie et les Pays-Bas pourraient être rattachés à cet ensemble) présentent au contraire une faible homogamie et une faible hypergamie étant donné leur niveau d'ouverture sociale et de libéralisme culturel. L'homogamie socioéconomique est la plus nettement associée à cet axe, qui contribue très fortement à son explication.

Quels sont les facteurs qui définissent cet axe ? Le haut du troisième axe est associé à un écart de niveau d'éducation entre les sexes relativement favorable aux femmes (*Écart F-H édu.*), de même que moins nettement à un écart relativement favorable en termes socioéconomiques (*Écart F-H socioéco.*), à une importance accordée à la prise de risque (*Stimulation*) et à la liberté de décision (*Autonomie*), mais aussi – quoique moins nettement – à des inégalités économiques plus prononcées (*Gini*). L'ensemble de ces traits peuvent être résumés par la dominante protestante (*Protestant*) des régions du Nord de l'Europe. Pour

finir, notons qu'une proportion élevée de personnes résidant en zone rurale (*Ruraux*) n'est pas en elle-même associée à un renforcement de l'homogamie, au contraire de la proportion d'actifs travaillant dans le secteur agricole (*Agriculture*).

Le bas du troisième axe est au contraire associé à un PIB par habitant élevé et à un fort lien entre niveau d'éducation et groupe socioéconomique (*Éduc.-socioéco.*). Il apparaît ainsi qu'à niveau d'ouverture sociale et de libéralisme culturel fixé, le niveau de développement économique renforce l'homogamie socioéconomique, ainsi que moins nettement l'homogamie d'éducation et l'hypergamie. Cet effet apparaît le plus clairement dans le cas des capitales, déjà évoqué au chapitre précédent, qui sont placées plus bas que les régions qui les entourent sur la troisième dimension : Londres (UKI), Madrid (ES3), Paris et Île-de-France (FR1) ; le pays-capitale Luxembourg (LU0) apparaît quant à lui clairement en position extrême. Ainsi, les régions les plus développées rejoignent les régions les plus pauvres sur cet axe secondaire (effet Guttman partiel).

Le troisième axe apporte donc des réponses aux questions théoriques posées plus haut. Si le développement économique est associé à une diminution de l'homogamie et de l'hypergamie, c'est principalement du fait de son lien avec le libéralisme culturel et l'ouverture sociale. Une fois ces dimensions contrôlées, il est associé à une plus forte homogamie socioéconomique. Il ne semble pas que l'on assiste à une augmentation de l'homogamie aux niveaux de développement les plus élevés, comme l'a observé une étude existante à propos de l'homogamie d'éducation (Domański et Przybysz, 2007) : en effet, développement, libéralisme culturel et ouverture sociale vont généralement de pair, comme on le voit à l'alignement de la plupart des régions sur le plan¹⁴. C'est plutôt la faible homogamie (au regard de leur PIB par habitant) des régions du Nord de l'Europe situées en haut du troisième axe qui fait exception à la règle européenne.

Parmi ces régions, un statut relativement favorable des femmes est associé à une plus faible hypergamie relative, mais aussi à une moindre homogamie, au-delà même de l'effet du libéralisme culturel et de l'ouverture sociale (premier axe). Mais l'exemple de ces régions montre qu'une faible homogamie n'est pas synonyme de faibles inégalités de revenus. Il semble que la combinaison de valeurs individualistes et d'inégalités économiques dans les

14 Dans des analyses non reprises ici, aucune relation curvilinéaire entre homogamie et PIB par habitant n'a pu être relevée.

pays à dominante protestante du Nord de l'Europe soit congruente avec une faible homogamie. Toutefois, une pauvreté fréquente (*Pauvreté rel., Privation mat.*) est positivement associée à l'homogamie et à l'hypergamie : nos hypothèses sur les risques associés à une mésalliance sont confirmées.

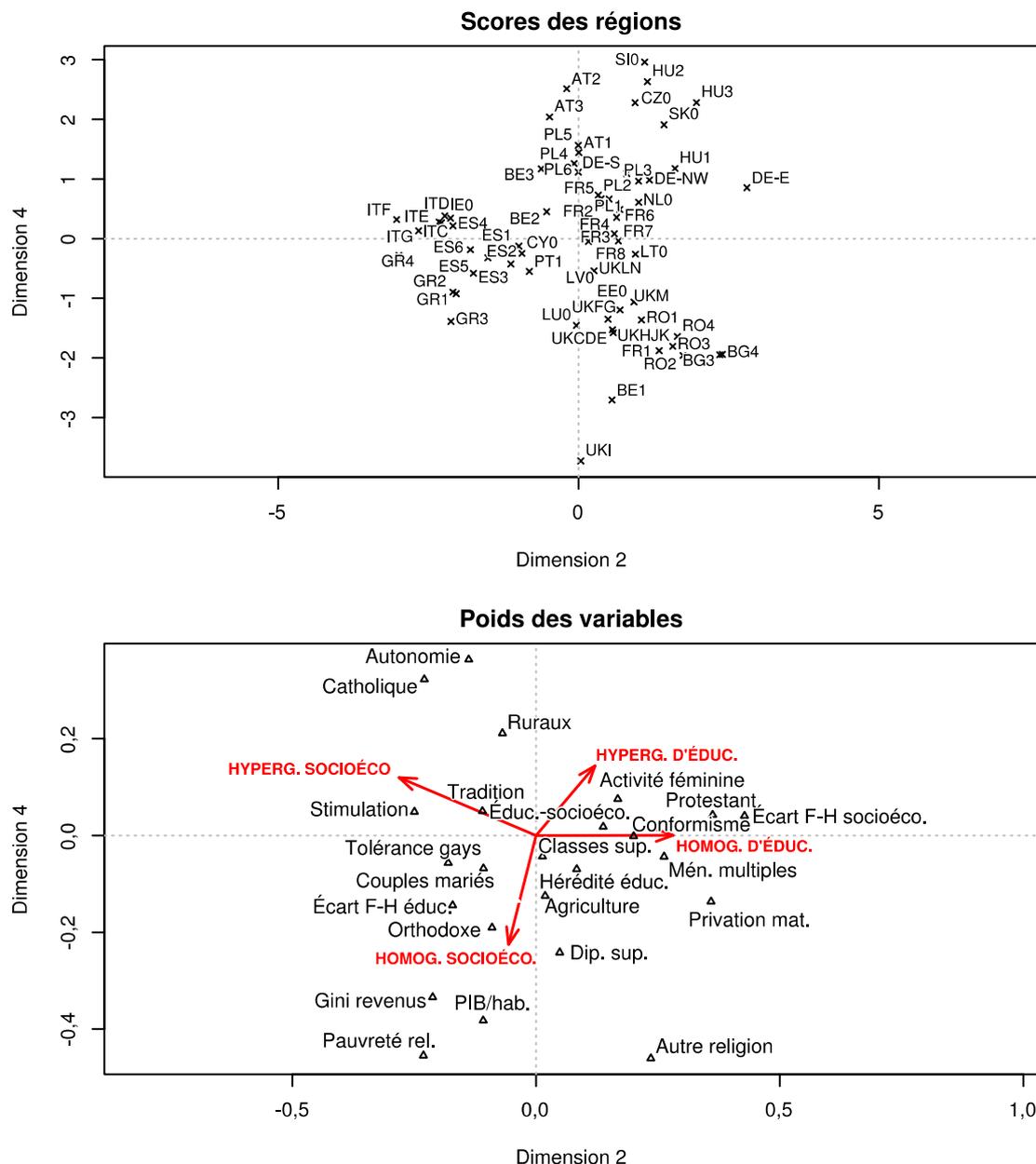
5.2 Des divergences : des aires géographiques aux structures sociales contrastées

Si les dimensions 1 et 3 représentaient à quelques nuances près des déterminants communs aux deux composantes de l'homogamie comme de l'hypergamie, les dimensions 2 et 4 (Figure 8.2) sont chacune l'occasion de mettre en évidence leurs oppositions. Ainsi, la deuxième dimension (horizontale) oppose composante éducative et composante socioéconomique du choix du conjoint, alors que la quatrième (verticale) oppose homogamie et hypergamie.

À position sur les autres axes fixée, la coordonnée sur le deuxième axe détermine le rapport entre intensité de l'homogamie et de l'hypergamie d'éducation (droite de l'axe) d'un côté, et socioéconomique (gauche de l'axe) de l'autre. La variable la plus déterminante dans la construction de l'axe est le degré de protestantisme (*Protestant*), associé à la droite de l'axe, et donc à une plus forte homogamie et hypergamie d'éducation. Il s'oppose au degré de catholicisme (*Catholique*) et de christianisme orthodoxe (*Orthodoxe*), dont la position à gauche de l'axe est cependant moins marquée.

Les régions de la droite de l'axe se caractérisent par des taux d'activité féminine (*Activité féminine*) relativement élevés, et plus encore par un écart de statut socioéconomique entre hommes et femmes (*Écart F-H socioéco.*) relativement favorable à ces dernières. À l'inverse, dans les régions de la gauche de l'axe, les femmes bénéficient d'une situation plus favorable dans le domaine éducatif (*Écart F-H éduc.*), mais ne convertissent pas cet avantage sur le marché du travail. Ainsi, il n'est pas étonnant de trouver dans le premier groupe de pays une plus forte hypergamie féminine en termes d'éducation, et dans le second une plus forte hypergamie en termes socioéconomiques – même si l'influence de la structure de la population est ici contrôlée.

Figure 8.2 : Les dimensions 2 et 4 de la régression PLS



Source : Enquête européenne sur les forces de travail 2011 ; Enquête sociale européenne, vagues 1 à 5 (2002-2010) ; Eurostat ; Standardized World Income Inequality Database.

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour le groupe socioéconomique) pour la mesure de l'homogamie et de l'hypergamie ; ensemble de la population de plus de 15 ans pour les autres variables (sauf exception, voir le corps du texte).

Cf. Tableau 7.5 (p. 454) pour la correspondance entre codes et noms des régions.

Cette opposition qui apparaît à première vue religieuse en recoupe en fait une autre, géographique. La régression PLS permet en effet d'avancer des interprétations que les variables ne permettent pas de mettre à l'épreuve, en revenant aux « noms propres » des régions projetées sur la seconde dimension (Figure 8.2). De fait, il apparaît que la moitié droite de l'axe contient sans exception toutes les régions des pays ayant connu un régime communiste, dont notamment la Hongrie, la République tchèque et la Slovaquie, majoritairement catholiques, ou encore la Bulgarie et la Roumanie, à forte dominante orthodoxe. Ces régions, et notamment celles occupant les positions les plus extrêmes à la droite de l'axe, portent la trace de la politique favorable à l'activité professionnelle des femmes de l'ère communiste.

Dans la moitié gauche de l'axe, associée à une homogamie et une hypergamie plus forte en termes socioéconomiques, sont regroupés des pays catholiques ou orthodoxes, mais qui ont surtout en commun d'appartenir à l'Europe du Sud, à l'exception de l'Irlande, ainsi que de certaines régions belges et autrichiennes dont la position sur l'axe est peu marquée. Ces régions présentaient déjà au chapitre précédent une structure de l'homogamie économique commune (Figure 7.4, p. 432) : les classes supérieures sont dans ces régions peu unifiées entre cadres d'un côté et professions libérales et intellectuelles de l'autre, alors que les classes populaires sont plus homogènes ; enfin, les petits indépendants présentent une faible endogamie. Alors que la présente analyse ne tient pas compte de la structure de l'homogamie, on voit que les autres caractéristiques de ces régions amènent ici encore à les regrouper. En plus des variables déjà commentées, elles présentent une homogamie et une hypergamie plus fortes en termes socioéconomiques qu'éducatifs.

Cette opposition s'étend à l'homogamie. À la droite de l'axe, se trouvent des régions dans lesquelles la proportion de diplômés du supérieur (*Dip. sup.*) est plus élevée, mais le lien entre niveau d'éducation des individus et celui de leur père (*Héritité éduc.*) plus étroit, de même que celui entre diplôme et groupe socioéconomique (*Éduc.-socioéco.*)¹⁵. L'éducation, étant socialement plus sélective et professionnellement plus rentable, y constitue donc un critère de choix du conjoint plus déterminant. Au contraire, dans les régions de la moitié gauche de l'axe, les inégalités économiques sont plus élevées (*Gini revenus*), mais le lien entre diplôme et groupe socioéconomique (et sans doute au-delà, avec le revenu) est plus

15 Nous avons vérifié que c'était encore le cas lorsque l'association parmi les hommes et celle parmi les femmes étaient distinguées.

distendu : l'éducation n'étant pas un critère de stratification aussi fondamental, l'homogamie socioéconomique l'emporte. Il est frappant de constater que la domination masculine épouse les contours de la stratification sociale, et s'exprime par les dimensions qui sont les plus significatives dans une société donnée.

Rappelons que cette opposition, présente sur le second axe, n'est cependant que secondaire : elle n'apparaît qu'entre régions présentant des degrés de libéralisme culturel, d'ouverture sociale et de développement similaires.

Enfin, ce n'est que sur le quatrième axe que des divergences entre homogamie et hypergamie apparaissent. Ce résultat est notable, et rejoint l'observation que nous venons de formuler : il signifie que les déterminants du choix du conjoint sont globalement les mêmes pour les deux phénomènes, les caractéristiques les plus distinctives étant valables aussi bien pour les dimensions symétriques entre hommes et femmes que pour les dimensions genrées. En outre, seule l'homogamie socioéconomique est concernée par cet axe, l'homogamie d'éducation occupant une position neutre.

Les déterminants de l'opposition entre les deux dimensions de l'hypergamie et l'homogamie socioéconomique sont loin d'être inattendus. En haut du quatrième axe, les régions à dominante catholique (*Catholique*), qui valorisent la liberté de décision (*Autonomie*), présentent une homogamie faible relativement à l'hypergamie. En bas de cet axe, les régions dans lesquelles les femmes occupent des positions relativement favorables du point de vue de l'éducation (*Écart F-H éduc.*), et secondairement de celui du statut socioéconomique (*Écart F-H socioéco.*), connaissent une hypergamie faible par rapport à l'homogamie socioéconomique. Ces deux dernières variables s'opposent en fait chacune de manière très nette à un type d'hypergamie : l'écart en termes d'éducation à l'hypergamie relative d'éducation, l'écart en termes socioéconomiques à l'hypergamie socioéconomique¹⁶. L'effet de la proportion de ruraux (*Ruraux*) est difficile à interpréter.

D'autres facteurs entrent cependant en jeu. Ce n'est que sur ce dernier axe que les différents indicateurs du niveau des inégalités se rejoignent, associés cette fois au niveau de développement (*PIB/hab.*) : de fortes inégalités de revenus (*Gini*), un taux élevé de personnes sous le seuil de pauvreté relative (*Pauvreté rel.*) et en situation de privation matérielle sévère

16 Le taux d'activité féminine est aussi négativement associé à l'hypergamie socioéconomique, même s'il est situé légèrement en haut de l'axe.

(*Privation mat.*) sont associées à une plus forte homogamie socioéconomique. Enfin, du point de vue religieux, les régions marquées par une forte homogamie socioéconomique comprennent le plus souvent de fortes minorités religieuses (*Autre religion*), caractéristique encore une fois associée aux capitales comme Londres (UKI) et Bruxelles (BE1), mais aussi à la Bulgarie du fait de la présence d'une minorité musulmane.

Le croisement des axes 2 et 4 dessine finalement une opposition en trois pôles. À gauche, les régions du Sud de l'Europe et l'Irlande, dont les caractéristiques ont déjà été commentées : l'homogamie et l'hypergamie y sont plus fortes en termes socioéconomiques que d'éducation. Les femmes y ont un niveau d'éducation élevé, mais sont peu actives sur le marché du travail, ce qui se traduit par une hypergamie socioéconomique élevée, au-delà même de l'influence de la structure de la population.

En haut à droite, les pays d'Europe centrale, au passé communiste ou non : Slovénie, Hongrie, République tchèque, Slovaquie, Autriche, Allemagne, Pays-Bas, Belgique. Ces régions se caractérisent par une hypergamie d'éducation particulièrement forte au regard de leur homogamie socioéconomique. Le niveau d'éducation des femmes y est peu élevé au regard de leur situation sur le marché du travail.

En bas à droite du plan, on retrouve certains pays du Nord de l'Europe (Estonie, Lettonie, Lituanie, Royaume-Uni), la Roumanie, la Bulgarie et les capitales. Ces régions ont en commun une hypergamie socioéconomique très faible, mais une homogamie relativement élevée tant en termes d'éducation que de groupe socioéconomique. L'hétérogénéité de ce regroupement, dont les régions sont par ailleurs très éloignées sur les autres dimensions, est cependant trop forte pour lui donner une interprétation claire.

6 Conclusion

Les analyses conduites dans ce dernier chapitre valident largement le cadre théorique élaboré en introduction de cette thèse et précisé dans ses trois parties. Tout d'abord, les théories de la modernisation trouvent une confirmation dans la première dimension déterminant à la fois le niveau de l'homogamie et de l'hypergamie dans les régions d'Europe : celle-ci correspond au degré d'ouverture sociale et de libéralisme culturel. Elle

explique à elle seule 70 % des variations de l'homogamie d'éducation, et 40 % de celles de l'homogamie socioéconomique.

Cette confirmation vient donc préciser ces théories : si la modernisation entraîne bien une moindre détermination sociale du choix du conjoint, ce n'est pas tant comme effet mécanique de l'augmentation du niveau de vie que par une augmentation de l'entrecroisement des appartenances, une valorisation de l'autonomie individuelle et un moindre contrôle social. Ces caractéristiques sont associées à une faible population agricole, un faible taux de pauvreté et de privation matérielle, une faible proportion de couples mariés et de ménages multiples, et une faible valorisation de la tradition et du conformisme.

Aucune tendance à l'augmentation de l'homogamie avec certains aspects de la modernisation n'a pu être relevée, à l'inverse des résultats de certains travaux existants (Domański et Przybysz, 2007). L'homogamie est certes légèrement plus élevée dans les grandes capitales que dans les régions qui les environnent, mais rien n'indique que cette particularité soit un signe avant-coureur d'une évolution qui devrait se généraliser. Nous faisons l'hypothèse qu'elle s'explique par une ségrégation socio-spatiale plus marquée dans les grandes métropoles, associée au fait que les marchés conjugaux sont suffisamment étendus pour que les individus trouvent facilement un conjoint dans leur propre groupe – des questions qui pourraient être approfondies à l'aide de données plus détaillées¹⁷.

De son côté, le lien entre homogamie et niveau global des inégalités de revenus (indice de Gini) n'est pas très marqué, en cohérente avec l'ambivalence théorique du lien entre conscience de classe et inégalités. Celles-ci apparaissent associées, dans les pays d'Europe du Nord (hors Scandinavie), souvent à dominante protestante, à une valorisation de l'autonomie individuelle et de la prise de risque qui semblent s'accorder avec une faible détermination sociale du choix du conjoint¹⁸. Néanmoins, sur d'autres dimensions, le niveau des inégalités et de la pauvreté est associé à une plus forte homogamie socioéconomique. Cette divergence, qui peut expliquer la difficulté à identifier un effet clair dans les travaux précédents portant sur l'homogamie ou sur la mobilité sociale (Breen, 2004). Elle fait aussi

17 Rappelons que le niveau d'ensemble des inégalités de revenus (indice de Gini) n'a pu être mesuré ici qu'au niveau national.

18 Nous ne pouvons, à l'aide de nos données purement transversales, exclure la possibilité que cette coïncidence ne soit que temporaire. Par exemple, les inégalités ont pu évoluer bien plus rapidement dans les États baltes, depuis la fin de l'Union soviétique, que les valeurs individualistes ou que l'homogamie – même si cette dernière semble avoir réagi assez rapidement dans d'autres anciens pays communistes (Katrňák, Kreidl et Fónadová, 2006 ; Katrňák, 2008).

écho à la faiblesse de l'influence de l'homogamie d'éducation sur la concentration des revenus entre ménages, déjà discutée au chapitre 3.

Le fait que les déterminants de l'homogamie et de l'hypergamie soient largement communs invalide clairement l'hypothèse selon laquelle l'augmentation de l'activité professionnelle des femmes serait à l'origine d'un renforcement de l'homogamie. Une position plus favorable des femmes en termes éducatifs et professionnels est ainsi associée à une moindre hypergamie comme à une moindre homogamie. Plus en détail, nous avons observé que la domination masculine suivait les lignes de la stratification sociale. Là où l'éducation est relativement peu sélective socialement et déconnectée du statut professionnel, comme dans le Sud de l'Europe et en Irlande, l'homogamie socioéconomique est plus forte, et l'homogamie d'éducation plus faible ; les femmes sont plus diplômées, et l'hypergamie d'éducation moindre, mais cette situation ne se traduit pas en termes professionnels, et l'hypergamie socioéconomique est forte. En revanche, là où le lien entre éducation et groupe socioéconomique est fort, et l'hérédité sociale en termes d'éducation élevée, l'homogamie s'exprime d'abord en termes d'éducation – sans que l'homogamie socioéconomique diminue pour autant ; les femmes occupent des positions professionnelles relativement élevées, et l'hypergamie socioéconomique est faible, mais l'hypergamie d'éducation est forte.

Les facteurs religieux offrent un résumé des oppositions entre régions. Les régions à fort degré de protestantisme présentent une homogamie et une hypergamie plus faibles ; l'homogamie d'éducation l'emporte sur l'homogamie socioéconomique. À l'inverse, les régions fortement orthodoxes présentent une homogamie et une hypergamie plus fortes, en particulier en termes socioéconomiques ; l'hypergamie apparaît relativement moins élevée que l'homogamie relativement aux autres régions. Finalement, les régions catholiques se distinguent avant tout par une forte hypergamie.

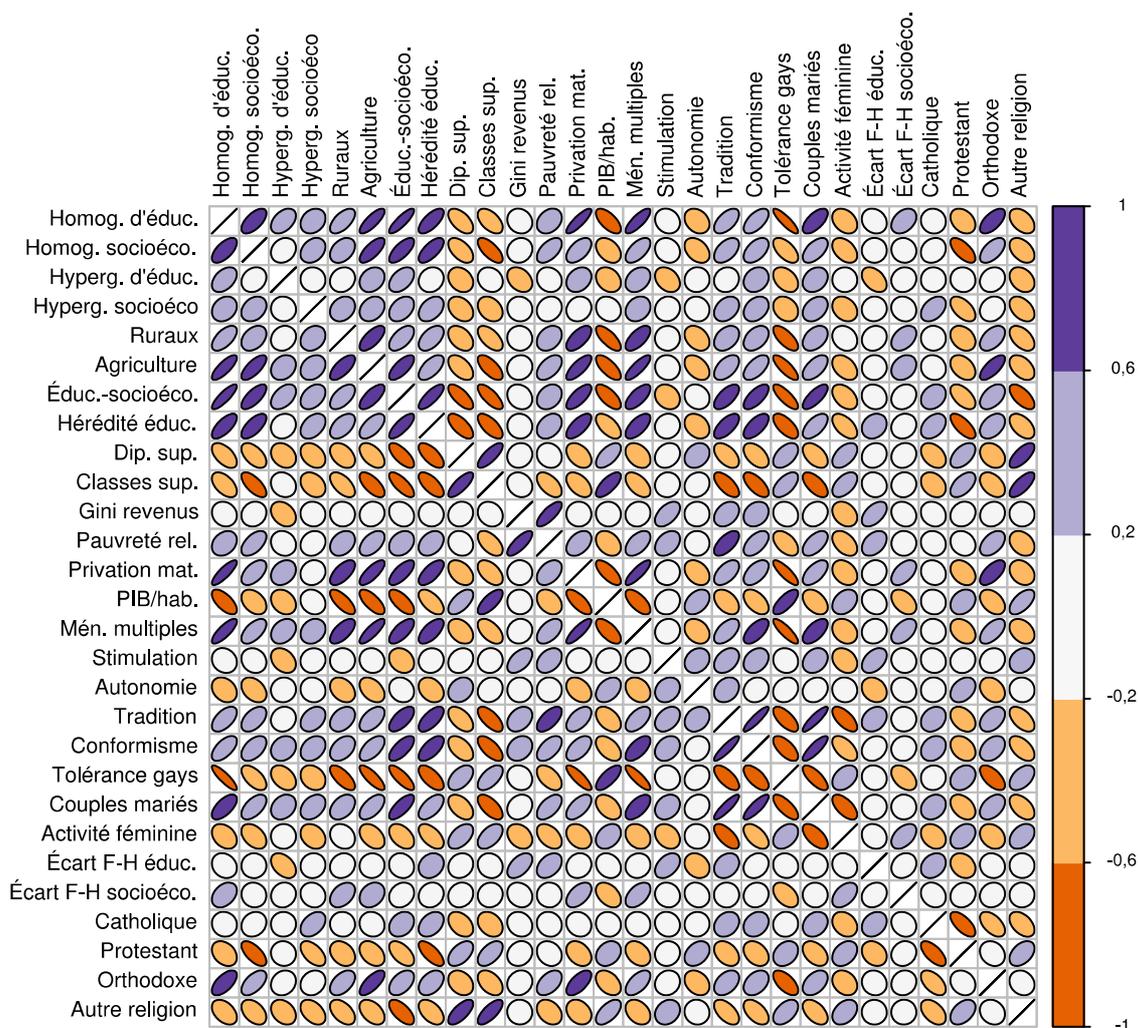
L'opposition entre Est et Ouest n'est pas apparue déterminante, celle entre Sud et Nord étant tout aussi nette. La seule particularité des anciens pays communistes est une situation relativement favorable des femmes sur le marché du travail, mais qui ne se traduit pas de manière systématique dans l'hypergamie.

En conclusion, il nous faut souligner que si l'explication de l'homogamie paraît satisfaisante, celle de l'hypergamie reste très imparfaite (entre 25 % et 30 % des variations).

Ce problème s'explique sans doute en partie par l'imperfection de la mesure, qui exige des effectifs très importants (*cf.* chapitre précédent). Cette étude pourrait être prolongée en recourant à des échantillons de plus grande taille, par exemple en retenant plusieurs années d'enquête. Néanmoins, cette explication nous semble partielle : la difficulté à expliquer les variations de l'hypergamie relative indique peut-être aussi une question théorique à approfondir.

7 Appendice : Valeurs des variables et corrélations deux à deux

Figure 8.3 : Corrélations entre variables prises deux à deux



Source : Enquête européenne sur les forces de travail 2011 ; Enquête sociale européenne, vagues 1 à 5 (2002-2010) ; Eurostat ; Standardized World Income Inequality Database.

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour le groupe socioéconomique) pour la mesure de l'homogamie et de l'hypergamie ; ensemble de la population de plus de 15 ans pour les autres variables (sauf exception, voir le corps du texte).

Les ellipses inclinées vers la droite correspondent à des corrélations positives, celles inclinées vers la gauche à des corrélations négatives, et les ellipses rondes à des corrélations nulles, recoupant ainsi l'information fournie par les couleurs.

Tableau 8.3 : Valeur des variables et effectifs de l'Enquête sociale européenne pour chaque région étudiée

	Effectifs ESS	Variables expliquées					Opport. de renc.	Appartenances entrecroisées					Inégalités		Niveau de vie	
		Homog. d'éduc	Homog. socioéco.	Hyperg. d'éduc.	Hyperg. socioéco		Ruraux	Agriculture	Éduc.-socioéco.	Héritité éduc.	Diplômés sup.	Classes sup.	Gini revenus	Pauvreté relative	Privation matér.	PIB/hab.
Unité		MAOR	GOR	AS		%	%	MAOR	%	%		%	%	%	%	
Moy.	2315	8,6	3,6	1,6	1,1	25	6,3	10,5	8,7	22	36	31	17	9,6	1,0	
Écart-type	2868	2,4	1,2	7,1	1,2	20	7,8	1,8	2,0	8	9	3	3	9,7	0,4	
AT1	3954	6,4	3,4	1,80	1,07	7	3,9	7,0	7,0	18	39	27	14	4,9	1,4	
AT2	2009	5,8	3,2	1,80	1,13	6,6	7,8	6,6	6,6	13	33	27	14	3,9	1,1	
AT3	3210	7,1	3,6	0,86	1,21	7,6	6,0	7,6	7,6	14	33	27	14	3,6	1,3	
BE1	578	5,6	4,3	0,59	0,87	7	0,2	7,0	7,0	35	45	26	15	17,9	2,4	
BE2	5530	6,9	3,4	0,70	1,07	9,9	1,4	9,9	9,9	27	41	26	15	2,3	1,2	
BE3	2831	4,5	3,0	1,60	1,41	8,1	2,1	8,1	8,1	26	39	26	15	8,6	0,9	
BG3	3364	21,6	3,6	0,34	0,98	12,9	9,3	12,9	12,9	15	24	30	20	51,6	0,3	
BG4	2700	51,2	3,6	0,36	0,68	18,4	8,1	18,4	18,4	22	32	30	20	45,8	0,5	
CY0	3293	6,6	4,2	1,10	1,12	10,9	4,4	10,9	10,9	30	32	28	16	11,3	1,0	
CZ0	8790	20,1	3,4	6,30	1,18	25,3	4,5	25,3	25,3	15	33	26	9	8,0	0,8	
DE-E	5449	7,9	2,8	2,40	0,88	8,2	2,8	8,2	8,2	28	40	28	14	5,0	0,9	
DE-NW	5853	7,8	3,2	2,00	1,00	10,2	1,8	10,2	10,2	21	40	28	14	5,0	1,2	
DE-S	3185	7,8	3,4	4,00	0,99	10,1	1,9	10,1	10,1	24	40	28	14	5,0	1,3	
EE0	6960	6,0	2,9	0,68	0,99	5,7	5,8	5,7	5,7	31	39	33	19	7,5	0,6	
ES1	1311	3,9	3,9	0,48	0,98	9,6	7,9	9,6	9,6	25	27	32	20	3,0	0,9	
ES2	912	3,7	3,3	1,30	1,52	10,5	3,3	10,5	10,5	32	31	32	20	2,0	1,3	
ES3	1276	6,9	4,0	0,64	1,42	9	0,7	9,0	9,0	36	40	32	20	4,1	1,4	
ES4	1251	3,3	3,9	0,74	1,10	12	8,4	12,0	12,0	22	25	32	20	2,3	0,9	
ES5	2602	3,8	4,5	0,91	0,95	10,1	2,9	10,1	10,1	25	26	32	20	4,1	1,1	
ES6	2032	4,0	4,5	1,80	1,03	11,9	11,4	11,9	11,9	21	23	32	20	6,1	0,8	

Libéralisme des mœurs et contrôle social						Statut des femmes				Degré de religiosité				
Mén. multiples	Stimulation	Autonomie	Conformisme	Tradition	Tolérance gays	Couples mariés	Activité féminine	Écart F-H éduc.	Écart F-H socioéco.	Catholique	Protestant	Orthodoxe	Autre	Unité
%		1-6			1-5	%	%	0-1	0-1			0-10		
3,4	3,0	4,8	3,9	4,3	3,8	72	75	0,47	0,36	3,8	1,1	0,6	0,5	Moyenne
4,4	0,2	0,2	0,5	0,4	0,5	14	8	0,05	0,05	2,2	1,5	1,7	0,5	Écart- type
3,6	3,4	5,0	3,4	4,1	3,9	70	80	0,39	0,38	4,9	0,3	0,2	0,6	AT1
7,4	3,3	4,9	3,4	4,2	3,9	59	78	0,35	0,36	5,2	0,3	0,0	0,3	AT2
3,8	3,3	4,9	3,4	4,1	3,9	64	78	0,37	0,34	5,5	0,2	0,0	0,2	AT3
0,0	3,2	5,0	3,7	4,1	4,1	93	68	0,49	0,37	4,0	0,5	0,3	2,7	BE1
0,0	3,1	4,8	3,9	4,3	4,2	82	77	0,49	0,36	5,8	0,1	0,0	0,7	BE2
0,0	3,1	5,0	3,6	4,1	4,1	89	69	0,50	0,39	6,2	0,1	0,1	0,7	BE3
11,6	3,3	4,4	4,3	4,8	3,5	87	77	0,50	0,45	0,0	0,1	3,5	1,2	BG3
9,1	3,3	4,5	4,4	4,9	3,5	87	83	0,53	0,49	0,1	0,0	3,8	1,1	BG4
5,4	3,3	4,9	4,0	5,1	3,3	87	77	0,50	0,31	0,0	0,0	6,8	0,0	CY0
1,8	2,9	4,7	4,2	4,1	3,7	74	80	0,41	0,34	4,9	0,4	0,0	0,3	CZ0
1,2	2,6	5,0	3,7	4,0	3,9	53	86	0,46	0,45	1,0	3,9	0,1	0,6	DE-E
0,6	2,8	5,0	3,6	4,1	4,0	72	78	0,38	0,35	2,4	2,7	0,1	0,5	DE-NW
0,7	2,8	5,0	3,5	4,1	3,9	74	82	0,35	0,34	3,4	1,9	0,1	0,5	DE-S
1,8	3,0	4,7	3,7	4,0	3,0	52	85	0,62	0,46	0,1	2,0	3,6	0,5	EE0
7,9	2,8	4,7	4,0	4,3	4,0	74	74	0,48	0,36	5,4	0,0	0,0	0,0	ES1
2,8	2,9	4,9	3,9	4,4	4,2	73	77	0,47	0,34	5,2	0,0	0,0	0,3	ES2
2,9	3,0	5,1	3,9	4,2	4,1	69	80	0,45	0,35	5,0	0,1	0,1	0,3	ES3
1,8	3,0	4,7	4,2	4,4	3,8	80	71	0,51	0,32	5,4	0,0	0,1	0,1	ES4
2,1	3,1	4,9	3,9	4,4	4,0	71	77	0,51	0,33	4,8	0,0	0,1	0,6	ES5
3,1	2,9	4,8	4,2	4,4	4,1	76	70	0,47	0,31	5,2	0,0	0,0	0,3	ES6

	Effectifs ESS	Variables expliquées					Opport. de renc.	Appartenances entrecroisées					Inégalités		Niveau de vie	
		Homog. d'éduc	Homog. socioéco.	Hyperg. d'éduc.	Hyperg. socioéco		Ruraux	Agriculture	Éduc.-socioéco.	Héritité éduc.	Diplômés sup.	Classes sup.	Gini revenus	Pauvreté relative	Privation matér.	PIB/hab.
Unité		MAOR	GOR	AS		%	%	MAOR	%	%		%	%	%	%	
FR1	1371	4,9	3,8	0,73	1,08	6,4	0,3	6,4	6,4	33	56	28	13	5,3	1,8	
FR2	1573	4,7	3,5	1,00	0,99	7,3	4,2	7,3	7,3	19	38	28	13	5,3	1,0	
FR3	669	6,4	3,5	0,72	1,06	7,2	2,0	7,2	7,2	20	42	28	13	5,3	0,9	
FR4	917	5,3	3,1	0,51	1,23	7,9	2,3	7,9	7,9	21	41	28	13	5,3	1,0	
FR5	1304	4,8	3,7	0,64	1,29	6,6	5,0	6,6	6,6	21	39	28	13	5,3	1,0	
FR6	1130	5,4	3,2	1,60	1,18	6,7	5,5	6,7	6,7	23	43	28	13	5,3	1,0	
FR7	1174	4,8	4,0	1,30	0,77	6,6	3,5	6,6	6,6	24	44	28	13	5,3	1,1	
FR8	958	4,1	2,9	1,40	1,22	6,9	3,1	6,9	6,9	22	41	28	13	5,3	1,0	
GR1	3267	7,2	3,5	0,67	1,21	13,5	18,9	13,5	13,5	18	25	33	20	12,1	0,8	
GR2	2154	9,3	3,4	0,76	1,49	14,7	22,5	14,7	14,7	14	21	33	20	14,1	0,8	
GR3	3361	12,0	4,3	1,30	1,51	14,4	1,2	14,4	14,4	25	35	33	20	9,5	1,2	
GR4	977	6,7	3,9	1,60	1,23	13,1	15,9	13,1	13,1	14	22	33	20	12,8	0,9	
HU1	1916	20,0	3,1	7,30	0,74	16,8	1,2	16,8	16,8	26	43	27	13	18,2	1,1	
HU2	2648	19,0	3,3	0,44	1,01	20,6	6,3	20,6	20,6	14	28	27	13	17,3	0,6	
HU3	3242	23,0	3,4	1,50	1,12	27,3	7,3	27,3	27,3	14	28	27	13	24,6	0,4	
IE0	10472	5,7	2,8	0,89	1,16	5	4,4	5,0	5,0	30	42	30	17	5,3	1,4	
ITC	591	7,4	4,3	0,78	1,22	10,7	2,2	10,7	10,7	12	32	33	19	3,3	1,3	
ITD	468	8,1	3,9	0,87	1,14	10,8	3,9	10,8	10,8	11	30	33	19	3,2	1,3	
ITE	521	8,0	3,9	1,30	1,18	11,7	3,0	11,7	11,7	13	30	33	19	4,8	1,2	
ITF	799	11,1	4,1	0,85	1,18	16,7	9,1	16,7	16,7	10	27	33	19	11,6	0,7	
ITG	357	10,6	4,1	1,40	1,06	16,3	8,7	16,3	16,3	9	26	33	19	14,7	0,7	
LT0	3679	11,1	3,8	0,70	1,39	14,5	9,3	14,5	14,5	25	39	34	20	20,4	0,6	
LU0	3187	8,7	5,5	1,20	1,10	15,1	1,3	15,1	15,1	29	54	27	14	1,0	2,7	
LV0	3940	8,0	3,1	5,30	1,39	9,6	8,9	9,6	9,6	23	36	35	23	27,3	0,5	
NL0	9741	4,2	2,7	0,79	1,06	6,8	2,8	6,8	6,8	26	45	27	10	1,9	1,4	

Mén. multiples	Libéralisme des mœurs et contrôle social						Statut des femmes			Degré de religiosité				Unité
	Stimulation	Autonomie	Conformisme	Tradition	Tolérance gays	Couples mariés	Activité féminine	Écart F-H éduc.	Écart F-H socioéco.	Catholique	Protestant	Orthodoxe	Autre	
%	1-6					1-5	%	%	0-1	0-1	0-10			
0,8	2,8	4,6	3,3	3,8	4,2	56	82	0,46	0,37	3,4	0,2	0,1	1,9	FR1
0,3	2,8	4,4	3,2	3,7	4,2	48	80	0,46	0,35	4,7	0,1	0,0	0,6	FR2
1,0	2,7	4,4	3,2	3,8	4,1	50	73	0,47	0,34	5,2	0,1	0,0	0,8	FR3
0,4	2,9	4,4	3,3	3,9	4,1	50	80	0,48	0,35	4,3	0,4	0,0	0,6	FR4
0,1	2,8	4,4	3,1	3,5	4,2	47	83	0,47	0,36	4,9	0,1	0,0	0,3	FR5
1,0	2,8	4,4	3,1	3,6	4,2	45	83	0,49	0,37	4,3	0,3	0,0	0,4	FR6
0,5	2,8	4,5	3,2	3,7	4,1	50	80	0,49	0,34	4,1	0,2	0,0	0,9	FR7
0,7	3,0	4,5	3,2	3,7	4,2	54	76	0,48	0,34	4,4	0,2	0,1	0,9	FR8
4,0	3,4	5,0	4,3	5,1	3,3	96	66	0,48	0,31	0,0	0,0	6,9	0,3	GR1
2,9	3,3	4,8	4,4	5,1	3,1	96	66	0,49	0,29	0,0	0,0	7,0	0,1	GR2
1,5	3,4	5,0	4,1	5,0	3,6	95	68	0,49	0,31	0,1	0,0	6,5	0,2	GR3
4,0	3,6	5,1	4,1	5,1	3,2	96	66	0,47	0,29	0,2	0,0	6,8	0,2	GR4
1,1	2,9	5,0	3,5	4,5	3,3	68	75	0,48	0,40	3,8	1,5	0,0	0,2	HU1
0,9	2,9	5,0	3,4	4,5	3,2	67	73	0,54	0,41	4,8	0,9	0,0	0,1	HU2
1,7	2,8	4,9	3,5	4,5	3,2	68	69	0,49	0,42	3,5	1,9	0,0	0,2	HU3
3,1	3,3	4,8	3,9	4,5	4,0	70	68	0,49	0,33	5,6	0,2	0,0	0,1	IE0
1,4	3,2	4,9	4,3	4,8	3,8	78	72	0,49	0,33	6,5	0,0	0,0	0,1	ITC
0,8	3,0	4,9	4,4	5,0	3,9	78	74	0,48	0,33	6,2	0,0	0,0	0,0	ITD
4,6	3,0	4,9	4,5	4,8	3,8	80	69	0,50	0,31	6,7	0,0	0,0	0,1	ITE
0,8	3,1	4,8	4,4	5,0	3,6	94	43	0,53	0,25	7,0	0,0	0,0	0,0	ITF
7,8	3,3	4,8	4,2	4,9	3,5	90	46	0,43	0,27	6,2	0,1	0,0	0,2	ITG
0,3	3,1	4,4	3,6	4,3	2,6	92	87	0,62	0,52	5,5	0,1	0,3	0,1	LT0
8,2	3,1	4,7	4,0	4,2	4,1	79	72	0,45	0,34	3,5	0,0	0,1	1,3	LU0
9,6	3,6	5,0	3,7	4,6	3,1	65	85	0,52	0,49	2,2	1,5	1,5	0,5	LV0
15,9	3,3	4,9	4,1	4,2	4,3	61	79	0,49	0,35	2,9	2,8	0,0	1,1	NL0

	Effectifs ESS	Variables expliquées					Opport. de renc.	Appartenances entrecroisées					Inégalités		Niveau de vie	
		Homog. d'éduc	Homog. socioéco.	Hyperg. d'éduc.	Hyperg. socioéco		Ruraux	Agriculture	Éduc.-socioéco.	Héritité éduc.	Diplômés sup.	Classes sup.	Gini revenus	Pauvreté relative	Privation matér.	PIB/hab.
Unité	MAOR	GOR	AS	%	%	MAOR	%	%	%	%	%	%	%	%	%	
PL1	1935	21,3	3,8	1,00	1,19	21,1	11,8	21,1	21,1	23	37	30	18	21,8	0,8	
PL2	1778	18,4	3,7	9,60	1,27	18,7	12,7	18,7	18,7	19	34	30	18	21,8	0,6	
PL3	1725	18,6	4,0	0,93	1,17	20,2	23,0	20,2	20,2	17	27	30	18	21,8	0,4	
PL4	1331	27,7	3,7	0,80	1,32	21,3	11,4	21,3	21,3	18	30	30	18	21,8	0,6	
PL5	817	26,3	3,5	1,30	1,06	18,6	9,5	18,6	18,6	17	32	30	18	21,8	0,6	
PL6	1331	24,9	3,6	6,70	1,08	23,1	10,7	23,1	23,1	17	31	30	18	21,8	0,5	
PT1	10302	8,9	4,5	1,30	1,06	15,9	9,8	15,9	15,9	14	25	35	18	9,3	0,8	
RO1	1030	66,7	6,2	0,44	1,39	29,1	23,1	29,1	29,1	10	21	32	23	23,9	0,4	
RO2	1293	221	7,8	2,40	1,03	38,0	41,1	38,0	38,0	9	18	32	23	39,2	0,3	
RO3	1113	49,3	6,4	17,0	1,07	36,9	22,4	36,9	36,9	16	28	32	23	35,7	0,6	
RO4	849	228	5,9	7,20	1,24	30,2	36,0	30,2	30,2	11	21	32	23	30,4	0,4	
SI0	7126	12,2	2,9	0,34	1,17	17,3	8,5	17,3	17,3	20	39	23	12	5,7	0,9	
SK0	6944	30,7	3,7	>999	0,97	16,5	4,4	16,5	16,5	15	32	26	12	14,7	0,7	
UKCDE	2802	3,8	2,8	0,75	0,90	5,1	1,1	5,1	5,1	28	41	35	18	4,4	1,0	
UKFG	1796	4,1	2,4	0,59	0,91	4,4	1,7	4,4	4,4	26	41	35	18	4,4	1,0	
UKHJK	3561	4,4	2,8	0,83	0,99	4,5	1,5	4,5	4,5	34	48	35	18	4,4	1,2	
UKI	968	4,3	3,2	0,66	0,94	4,2	0,2	4,2	4,2	46	56	35	18	4,4	2,0	
UKLN	991	3,7	2,9	1,50	0,76	5	2,3	5,0	5,0	28	42	35	18	4,4	0,9	
UKM	999	3,8	3,1	0,64	0,80	4,9	3,3	4,9	4,9	36	43	35	18	4,4	1,1	

Source : Enquête européenne sur les forces de travail 2011 ; Enquête sociale européenne, vagues 1 à 5 (2002-2010) ; Eurostat ; Standardized World Income Inequality Database ; se référer au corps du texte pour plus de détails.

Champ : couples cohabitants dans lesquels l'un des conjoints est âgé de 20 à 39 ans (et les deux conjoints ont travaillé dans les huit années précédant l'enquête pour le groupe socioéconomique) pour la mesure de l'homogamie et de l'hypergamie ; ensemble de la population de plus de 15 ans pour les autres variables (sauf exception, voir le corps du texte).

Cf. Tableau 7.5 (p. 454) pour la correspondance entre codes et noms des régions.

Libéralisme des mœurs et contrôle social						Statut des femmes				Degré de religiosité				Unité
Mén. multiples	Stimulation	Autonomie	Conformisme	Tradition	Tolérance gays	Couples mariés	Activité féminine	Écart F-H éduc.	Écart F-H socioéco.	Catholique	Protestant	Orthodoxe	Autre	
%		1-6			1-5	%	%	0-1	0-1			0-10		
11,7	3,0	4,9	4,5	4,8	3,3	91	77	0,52	0,44	6,5	0,0	0,0	0,1	PL1
7,6	3,1	4,8	4,6	4,9	3,1	92	71	0,49	0,41	6,8	0,0	0,0	0,1	PL2
7,3	3,0	4,8	4,6	4,9	2,9	96	76	0,50	0,47	6,8	0,0	0,2	0,0	PL3
4,7	3,3	4,8	4,5	4,8	3,2	88	69	0,50	0,40	6,6	0,0	0,0	0,1	PL4
12,3	3,1	5,0	4,6	4,8	3,3	88	72	0,54	0,43	6,6	0,0	0,0	0,1	PL5
12,8	3,1	4,7	4,5	4,8	3,1	90	68	0,43	0,40	6,8	0,0	0,0	0,1	PL6
17,3	3,0	4,4	3,8	4,3	3,7	75	81	0,43	0,38	5,9	0,1	0,0	0,2	PT1
19,3	3,1	4,4	4,4	4,7	2,7	92	66	0,42	0,38	0,9	0,9	4,6	0,4	RO1
3,6	2,8	4,3	4,6	4,9	2,5	92	67	0,40	0,40	0,2	0,1	6,4	0,1	RO2
8,6	3,2	4,5	4,5	4,7	2,7	89	68	0,47	0,40	0,1	0,0	6,4	0,0	RO3
1,8	2,8	4,4	4,7	4,9	2,7	94	66	0,45	0,37	0,4	0,3	6,2	0,1	RO4
2,0	3,3	5,0	3,9	4,4	3,3	60	80	0,49	0,49	5,4	0,1	0,1	0,2	SI0
1,2	3,0	4,6	4,4	4,5	3,1	89	78	0,48	0,38	5,8	0,6	0,1	0,5	SK0
3,4	3,3	4,8	3,7	4,1	4,0	57	77	0,50	0,38	1,2	3,4	0,0	1,3	UKCDE
1,0	3,3	4,8	3,7	4,1	3,9	62	77	0,48	0,36	0,8	3,8	0,0	1,1	UKFG
1,3	3,2	4,8	3,6	4,2	4,0	65	80	0,50	0,34	0,8	4,2	0,0	0,8	UKHJK
3,4	3,6	4,9	3,6	4,2	3,9	72	74	0,50	0,34	1,7	2,2	0,1	2,2	UKI
1,0	3,2	4,8	3,8	4,2	3,8	65	76	0,50	0,38	1,5	3,9	0,0	0,4	UKLN
1,3	3,2	4,8	3,7	4,1	3,9	60	78	0,50	0,41	1,5	4,0	0,0	0,6	UKM

Les moyennes et écarts-types sont pondérés par la proportion de chaque région dans la population totale, à l'exception des effectifs dans l'Enquête sociale européenne (ESS). La moyenne et l'écart-type géométrique sont présentés pour les variables mesurées par le MAOR.

Dans la ligne « Unité », « % » indique une proportion exprimée en pourcentages, « MAOR » l'*odds ratio* absolu moyen, et « GOR AS » l'*odds ratio* généralisé anti-symétrique. Le PIB est mesuré en référence à la moyenne pour toutes les régions (valeur 1).

CONCLUSION

Le choix du conjoint est un puissant révélateur des évolutions de la stratification sociale, des rôles de genre et des modèles familiaux. Quelque peu délaissée en France suite au constat d'une certaine persistance de l'homogamie, cette question à la croisée de la sociologie de la stratification et de celle de la famille semble pourtant avoir beaucoup à nous apprendre sur l'évolution des sociétés contemporaines, et sur les tendances qu'elles suivent à plus long terme. À l'encontre d'un thème fréquemment repris dans la sociologie française, mais en cohérence avec nombre de travaux existants, nous avons mis en évidence un affaiblissement notable de l'homogamie avec la modernisation, qui se retrouve aussi bien dans les évolutions temporelles constatées en France depuis quarante ans qu'entre régions d'Europe. En ce qui concerne les asymétries entre hommes et femmes dans le choix du conjoint, la tendance à l'hypergamie féminine n'est pas apparue comme un phénomène absolument systématique en Europe, et son affaiblissement – voire son inversion, du point de vue du diplôme – en France au cours du temps est lui aussi marqué.

Quatre résultats principaux nous semblent devoir être soulignés : l'influence des contraintes structurelles ; les effets de la modernisation et de l'individualisation ; les conséquences de l'élévation du statut des femmes ; la nature et l'évolution des critères déterminant le choix du conjoint. Nous concluons par une évaluation des limites de nos analyses et par une présentation de pistes de recherche qui nous semblent à privilégier.

1 L'influence des contraintes structurelles

Suivant une interprétation inspirée de Georg Simmel, les contacts entre groupes sociaux sont largement déterminés par la structure de la population, et notamment par l'entrecroisement des appartenances. Cette lecture est apparue très féconde pour notre étude : tout au long de cette thèse, la structure du marché conjugal s'est révélée

déterminante pour le choix du conjoint. Ce mécanisme découle bien sûr avant tout d'une nécessité mathématique qui contraint fortement les possibilités de mise en couple. Néanmoins, il n'agit pas seulement, de manière relativement triviale, au niveau de l'ensemble d'une population (ici, nationale ou régionale). Les individus évoluent en effet sur de multiples marchés conjugaux locaux formés par la juxtaposition ou l'intersection de différents critères : comme le soulignait Simmel, âge, position sociale, lieu de résidence, lieu d'études ou d'activité professionnelle se conjuguent pour restreindre, ou du moins orienter, le choix du conjoint.

Ainsi, nous avons vu au chapitre 5 que la combinaison de l'hypergamie féminine et des normes concernant l'écart d'âge était à l'origine d'une situation défavorable sur le marché conjugal pour les femmes les plus diplômées, expliquant leur taux de célibat élevé. De même, le chapitre 6 a montré que l'âge et la scolarisation influençaient le choix du conjoint en termes de diplôme, et, ce, de manière différente selon le sexe et la cohorte de naissance, du fait de leur effet sur la structure du marché conjugal fréquenté par ou accessible aux individus. Enfin, l'entrecroisement des appartenances est apparu au chapitre 8 comme un déterminant majeur réduisant l'homogamie, du simple fait qu'il rend plus difficile à réaliser la correspondance parfaite entre caractéristiques des conjoints – notamment diplôme et groupe socioéconomique.

Le mécanisme relativement simple liant contraintes structurelles et choix du conjoint prend ainsi des formes complexes et variées selon les contextes. En tant que déterminant fondé sur des contraintes objectives tenant à la composition et à la distribution des groupes sociaux, il est en outre relativement aisément accessible à l'analyse sociologique – en particulier quantitative. Il apparaît donc comme une entrée privilégiée pour l'étude des relations entre groupes sociaux, et que l'on ne saurait en tout cas négliger.

2 Modernisation, individualisation, libéralisme culturel et ouverture sociale

Le libéralisme culturel et l'ouverture sociale sont clairement apparus au chapitre 8 comme les déterminants principaux du degré de détermination sociale du choix du conjoint – tant de l'homogamie que de l'hypergamie. Ces deux phénomènes sont difficilement

séparables tant dans les évolutions temporelles récentes en France qu'entre régions européennes. Pour cela, il semble qu'ils doivent être rattachés au mouvement plus large de modernisation et d'individualisation des sociétés. De fait, l'essor du sentiment amoureux comme seul déterminant légitime du choix du conjoint est lui aussi un phénomène séculaire qui a partie liée avec ces processus. Il est frappant de constater que l'on n'observe, du moins dans le champ temporel et géographique retenu, aucune divergence entre les effets des variations dans les structures sociales – plus ou moins ouvertes – et dans les modèles conjugaux – plus ou moins libéraux – sur le choix du conjoint.

Ce résultat soutient l'interprétation suivant laquelle le choix du conjoint est à la fois un symptôme et un déterminant de la stratification sociale. Il ne semble guère possible de maintenir une forte étanchéité et une forte identité des groupes sociaux dans un contexte d'alliances libres des individus – conformément à ce qu'avait déjà relevé Max Weber. Inversement, l'affaiblissement du contrôle social sur le choix du conjoint, comme les autres évolutions contemporaines des modèles familiaux, n'est envisageable que dans une situation de relative prospérité économique et de faible dépendance à l'égard des solidarités familiales. L'hypothèse de l'amour comme luxe que peuvent se permettre les sociétés ayant atteint un certain niveau de confort matériel (Smits, Ultee et Lammers, 1998) est conforme à ces observations, encore qu'il soit plus probable qu'il s'agisse d'une évolution progressive plutôt que d'un effet de palier.

Pourtant, si le lien entre hétérogamie, mobilité sociale et entrecroisement des appartenances est assez net, justifiant le recours à la notion d'ouverture sociale, le lien entre cette dernière dimension et les inégalités économiques est tout sauf évident. Les seules composantes des inégalités économiques qui soient clairement associées à l'homogamie sont la proportion de la population sous le seuil de pauvreté et celle en situation de privation matérielle sévère (chapitre 8). Ces résultats, conformes à ceux des études de la mobilité sociale (Breen, 2004), sont aussi cohérents avec le constat d'une augmentation des inégalités économiques dans les pays développés (OCDE, 2011) en dépit d'un affaiblissement de l'hérédité sociale. L'ouverture sociale pourrait ainsi être plus liée à l'égalité des chances et à la liberté individuelle qu'à l'égalité des positions. De fait, le déterminant principal d'une faible homogamie est le libéralisme culturel, et l'importance accordée à l'autonomie individuelle est un facteur secondaire important (chapitre 7). L'individualisation est bien le

trait commun au libéralisme culturel et à l'entrecroisement des appartenances, et donc le facteur principal de l'homogamie et de l'hypergamie.

3 Statut des femmes, choix du conjoint et inégalités

Outre l'effet général de la modernisation, une question théorique récurrente dans la littérature concerne l'effet de l'augmentation de l'activité des femmes sur l'homogamie (Mare, 1991 ; Blossfeld et Timm, 2003), et plus largement sur les inégalités. Ce raisonnement se fonde sur le principe selon lequel le retrait des femmes du marché du travail les mettrait en quelque sorte hors-jeu en ce qui concerne les questions de statut social. Nous avons vu que cette vision est très partielle : en se limitant à une définition strictement économique du statut social et des enjeux qui s'y rattachent, cette hypothèse manque l'un des aspects centraux de la stratification sociale.

Ainsi, nos analyses ont montré qu'en France au cours des quarante dernières années (chapitre 2) comme dans l'Europe actuelle (chapitre 8), l'augmentation de l'activité des femmes, et plus généralement l'amélioration de leur statut en termes éducatifs et professionnels, va de pair avec une diminution de l'homogamie. On peut expliquer ce résultat par le fait que la plus grande indépendance des femmes en termes économiques leur permet d'attribuer moins d'importance au statut social de leur conjoint, au profit de critères moins directement liés à la stratification sociale – ce que les hypothèses se fondant principalement sur l'évolution des préférences masculines avaient manqué.

Par ailleurs, notre étude de l'évolution du lien entre salaires des conjoints en France depuis 1990 (chapitre 3) a montré que l'augmentation de l'activité féminine et de la contribution des femmes au total des salaires des couples avait entraîné une diminution des inégalités entre couples. L'enseignement relativement contre-intuitif de ce résultat est que le passage d'un modèle dans lequel chaque couple touche un seul salaire à un autre dans lequel chaque couple en touche deux ne peut entraîner d'augmentation des inégalités que si les couples sont extrêmement homogènes du point de vue des salaires des conjoints. Or, l'association de deux individus, aussi semblables soient-ils, revient très rarement à unir deux clones : en conséquence, les inégalités entre couples diminuent généralement suite à cette

évolution. Seule une augmentation de l'activité féminine nettement supérieure à la moyenne parmi les classes supérieures pourrait contrer ce phénomène.

De manière plus attendue, l'augmentation du niveau d'éducation des femmes et de leur statut professionnel tend à diminuer l'hypergamie féminine selon ces dimensions. Cet effet est pour une part mathématique, par le biais des contraintes structurelles, mais il persiste au-delà de l'évolution de la composition de la population (hypergamie relative).

Dans le cas de la France, nous avons observé aux chapitres 4 et 5 que l'hypergamie relative d'éducation s'était inversée au moment même où les femmes étaient devenues plus diplômées que les hommes, sans opposer la moindre inertie. La norme d'hypergamie féminine est ainsi apparue bien faible : elle semble avoir tenu principalement le statut de modèle qu'acquiert la situation statistiquement la plus fréquente dans la population, indiquant une porosité entre les deux sens du mot *normal*. En outre, du point de vue de la profession, dont on pourrait penser qu'elle touche au cœur des rôles masculins, l'hypergamie relative a presque disparu au cours des quarante dernières années, alors que l'égalité professionnelle entre hommes et femmes est loin d'être atteinte. Ces résultats indiquent une évolution des normes conjugales vers l'idéal d'égalité entre conjoints plus rapide que celle du statut professionnel des femmes en général. Ce constat peut être nuancé en notant le maintien de l'hypergamie en termes de salaire parmi les couples bi-actifs, qui contraste avec le net développement de l'activité féminine et des salaires correspondants (chapitre 3).

Le lien entre statuts éducatif et professionnel des femmes et diminution de l'hypergamie a aussi été relevé en Europe (chapitre 8). La diversité des régions européennes nous a cependant permis d'identifier une nuance à cette règle, l'hypergamie étant plus élevée en termes socioéconomiques dans les pays d'Europe du Sud, et en termes d'éducation en Europe centrale. Instructive, cette opposition montre que la domination masculine s'exprime selon les critères qui apparaissent les plus déterminants pour l'homogamie, et plus largement pour la stratification d'une société. Elle peut être mise en regard avec les évolutions temporelles relevées dans le cas de la France (chapitres 3 et 4), où l'hypergamie de diplôme s'est inversée alors que l'hypergamie de classe sociale et surtout celle de salaire ont mieux résisté. L'hypothèse, avancée à propos de la France, selon laquelle le cœur des

rôles de genre tiendrait aux aspects économiques, la dimension culturelle n'étant que secondaire, ne semble pas valable pour tous les pays européens – et notamment pas pour ceux d'Europe centrale, dont les anciennes démocraties populaires qui ont connu une politique d'incitation au travail des femmes sous l'ère socialiste.

Dans l'ensemble, il est apparu clairement que la hausse du statut des femmes, intimement liée au libéralisme culturel, est associée à un affaiblissement aussi bien de l'hypergamie que de l'homogamie.

4 L'importance relative des critères de l'homogamie

Contrairement encore à une idée fréquemment reprise dans la littérature, nous n'avons pas observé de déplacement très net des critères de l'homogamie, que ce soit des caractéristiques héritées vers les caractéristiques acquises, ou des critères économiques vers les critères culturels. Ainsi, dans le cas de la France (chapitre 2), l'homogamie d'origine sociale n'a pas diminué plus rapidement que celle de diplôme ou de classe sociale. Tout juste avons-nous pu déceler une tendance à un affaiblissement légèrement plus rapide de l'homogamie de classe sociale par rapport à celle de diplôme depuis les années 1980 ; mais ce mouvement est particulièrement modéré, au regard des affirmations qui supposent l'existence d'une tendance majeure de long terme.

Aucun indice d'un déplacement des critères de l'homogamie n'est non plus apparu dans la comparaison des régions d'Europe (chapitres 7 et 8). Si une opposition entre les structures sociales accordant une plus grande place au diplôme et celles qui donnent plus d'importance au groupe socioéconomique est apparue, cette dimension ne paraît pas renvoyer à des degrés différents de modernisation, mais plutôt à la spécificité d'aires géographiques : contrairement au reste du continent, le diplôme est un déterminant relativement secondaire de la stratification dans l'Europe du Sud. De manière générale, nous avons observé que l'homogamie d'éducation est nettement supérieure à l'homogamie socioéconomique, et que les intensités respectives de ces deux associations ne sont similaires que lorsque l'homogamie d'éducation est faible – comme dans le cas de la France.

L'absence d'évolution systématique de l'importance relative des différentes dimensions de l'homogamie pourrait être attribuée à l'utilisation d'un indicateur peu adapté des déterminants économiques du choix du conjoint : comme nous l'avons défendu, la profession identifie tout autant des univers culturels que le diplôme. L'hypothèse d'une perte d'importance des aspects économiques devrait alors être testée à partir de l'étude de l'association entre revenus ou entre patrimoines des conjoints plutôt qu'à partir de l'homogamie socioéconomique. Nos observations concernant la France (chapitre 3) indiquent cependant que l'association entre salaires des conjoints parmi les couples bi-actifs a diminué à un rythme relativement faible, et qui n'apparaît en tout cas pas supérieur à celui de la baisse de l'homogamie de classe sociale ou de diplôme.

Au total, l'hypothèse d'un déplacement des critères de choix du conjoint dans un sens ou dans un autre apparaît donc peu soutenue par nos résultats. En outre, les travaux portant sur plusieurs composantes du choix du conjoint sont particulièrement rares : des analyses supplémentaires, en particulier concernant l'homogamie d'origine sociale – que nous n'avons pu étudier ici –, semblent requis.

5 Limites et perspectives

Si, dans l'ensemble, le cadre théorique développé en introduction à partir des théories de la modernisation et de la littérature existante s'est avéré remarquablement adapté à l'analyse du choix du conjoint dans la France des dernières décennies et dans l'Europe actuelle, il n'en demeure pas moins qu'il n'est pas en accord avec l'ensemble des résultats de la littérature. En effet, si la modernisation est toujours associée à un affaiblissement de l'homogamie, alors on s'attendrait à observer le même mouvement au cours du temps dans la plupart des pays – sauf exception justifiant une adaptation d'ordre secondaire de la théorie. Or, comme nous l'avons noté au chapitre 1, de nombreux travaux ont observé un renforcement de l'homogamie dans plusieurs pays, le cas le plus documenté étant celui des États-Unis. Plusieurs pistes permettant d'expliquer ce décalage peuvent être avancées.

La première, déjà proposée dans la littérature (Domański et Przybysz, 2007), serait que la courbe en U inversé liant homogamie et niveau de développement (Smits, Ultee et Lammers, 1998) se prolonge par une nouvelle phase d'augmentation de l'homogamie aux stades les

plus avancés du développement. Cependant, cette hypothèse ne semble guère correspondre à nos observations, dans la mesure où l'Europe comprend des régions parmi les plus développées, et que nous n'avons pu observer aucun mouvement de ce type – à l'exception notable de la plus forte homogamie des grandes capitales, dont il semble relativement risqué de considérer qu'elles anticipent les évolutions à venir des autres régions d'Europe.

Une seconde possibilité tiendrait à des différences dans la manière de mesurer l'homogamie. La question de la prise en compte ou non des couples cohabitants non mariés est évidemment essentielle, dans la mesure notamment où le niveau de libéralisme culturel est négativement associé à la fois à la proportion de couples mariés et à l'homogamie. Puisque les données de l'Enquête européenne sur les forces de travail contiennent les informations nécessaires, une analyse comparant les résultats calculés sur les seuls couples mariés, puis sur l'ensemble des couples, constituerait un apport intéressant pour la compréhension des divergences constatées dans la littérature, et permettrait de préciser le lien entre l'évolution des modèles conjugaux et le choix du conjoint.

Par ailleurs, une classification des diplômes plus fine que celle, en seulement trois niveaux, que nous avons retenue, pourrait éventuellement faire apparaître des distinctions restées invisibles dans notre étude. Cependant, des analyses utilisant une classification plus détaillée, restreintes aux pays qui proposent des données adéquates, fournissent des résultats très similaires à ceux que nous avons présentés. Cette piste ne semble donc pas suffisante pour expliquer les divergences entre travaux.

Une troisième possibilité, sociologiquement plus riche, impliquerait que la grande inertie des phénomènes étudiés introduit un décalage entre les évolutions temporelles au sein d'une société donnée, et les variations entre pays à un instant donné. En effet, confirmant un constat fait par les études de la mobilité sociale (Breen, 2004), les évolutions temporelles relevées dans le cas de la France nous sont apparues relativement limitées au regard des différences entre pays. Ce phénomène se retrouve aussi dans le cas de l'Irlande (Halpin et Chan, 2003) : l'augmentation actuelle de l'homogamie d'éducation relevée dans ce pays ne se reflète en rien dans notre analyse transversale, où il figure malgré tout parmi les moins homogames (chapitre 7). Il est possible qu'en dessous d'un certain niveau d'homogamie, des facteurs jusque-là secondaires, comme l'effet de la massification scolaire

sur les rencontres (chapitre 6) prennent le dessus sur les effets plus massifs de la modernisation. C'est en effet la diminution du temps écoulé entre études et mise en couple qu'invoquent les auteurs pour expliquer les évolutions observées dans le cas irlandais. Cette hypothèse expliquerait que l'homogamie puisse suivre des tendances apparemment erratiques parmi des pays pourtant similaires : parmi les pays à l'homogamie la plus faible, les variations temporelles seraient d'ampleur généralement limitée et dépendraient fréquemment de facteurs secondaires, plutôt que du niveau de modernisation.

Que cette piste d'explication soit correcte ou non, cette remarque plaide pour le recours à des indices qui fournissent une mesure *absolue* de l'intensité de l'association, et qui puissent être appliqués de manière uniforme quelle que soit la modélisation retenue – caractéristiques que possèdent le coefficient d'association intrinsèque et l'*odds ratio* absolu moyen que nous avons proposés (Annexe B). De tels indices permettraient de comparer aisément le *niveau*, et non seulement la tendance, de l'homogamie mesurés par les multiples études portant sur un ou une série de pays, ce qui améliorerait notre compréhension d'ensemble des variations de ce phénomène. En effet, se focaliser sur les évolutions temporelles sans pouvoir relativiser leur ampleur au regard des écarts entre pays mesurés par les travaux précédents n'apparaît pas comme le meilleur moyen d'élaborer une théorie globale des variations de l'homogamie.

La voie la plus directe pour approfondir cette question consistera cependant à mobiliser dans une seule analyse des données concernant plusieurs sociétés à des périodes suffisamment éloignées les unes des autres pour que les variations temporelles soient importantes. Les données de l'Enquête européenne sur les forces de travail, utilisées ici, permettraient de réaliser une telle étude à partir des années 1980, avec cependant comme limite importante que le nombre de pays couverts diminue nettement à mesure que l'on remonte dans le temps. Pour couvrir un champ géographique plus large, les travaux à venir devront donc probablement se fonder sur la combinaison de plusieurs enquêtes internationales, dont par exemple, outre l'Enquête européenne sur les forces de travail et l'Enquête sociale européenne, les recensements nationaux ou encore l'Étude luxembourgeoise sur les revenus (*Luxembourg Income Study*).

ANNEXES

ANNEXE A

LE CODAGE DE LA CLASSE ERIKSON-GOLDTHORPE-PORTOCARERO (EGP) DANS LES ENQUÊTES EMPLOI

Les tableaux suivants détaillent la manière dont est réalisé le classement des individus dans les différentes classes Erikson-Goldthorpe-Portocarero (EGP) pour les cinq séries d'enquête, à partir de la CSP ou de la PCS à deux chiffres, du statut de salarié ou d'indépendant, du nombre de salariés, de la qualification, de la profession détaillée et du secteur d'activité. La plupart des cas sont traités uniquement à partir de la CSP ou de la PCS (seconde colonne) ; les autres variables sont utilisées pour traiter les exceptions à cette règle (troisième colonne). Les variables et modalités correspondantes sont précisées dans le tableau A.5.

Le classement des individus selon leur classe d'origine (profession du père, ou mère ou tuteur au moment où le répondant a terminé ses études) suit le même principe, mais de manière simplifiée. Dans ce cas, seule la CSP/PCS et le statut (indépendant ou aide familial plutôt que salarié) sont connus ; les autres variables sont ignorées (tableau A.6). Le nombre de salariés étant renseigné seulement à partir de 1990, nous l'avons ignoré pour toutes les séries de manière à assurer leur cohérence : les catégories IVa et IVb ont été fusionnées en conséquence.

On pourra se référer au corps du texte (chapitre 1, p. 72) pour plus d'explications concernant les recodages.

Tableau A.1 : Codage de la classe EGP, 1969-1974 (Code des métiers de 1968)

Classe EGP	CSP	Exceptions
I	30, 32, 33, 34	-
II	41, 42, 80, 81, 44	9195 (Journaliste) en I 8374, 8391, 8392, 8393, 8394, 8396, 8397, 8398 (Acheteur, Chef des ventes/de rayon, Pompiste, Vendeur, Représentant de commerce, Voyageur de commerce, Commis d'agent de change, Démarcheur) en IIIa
IIIa	51	8180, 8572, 8679 (Standardiste/Télégraphiste, Gardien, Agent de service) en IIIb
IIIb	53, 70, 71, 72	4675 (Boucher) 8871 (Aide-infirmier) en IIIa 6571, 6572, 6573, 6593 (Chauffeurs) en VIIa 72 indépendant avec salariés en IVa 72 indépendant sans salariés en IVb
IVa	21, 22, 26, 27	sans salariés en IVb
IVb	21, 22, 26, 27	avec salariés en IVa
IVc	00, 23	-
V	43, 60	7571, 7572, 7576, 7594 (Dessinateurs), 7693, 7694, 7696, 7697, 8891 (Analyste programmeur, Analyste, Programmeur, Arpenteur/Topographe, Géomètre) en II 43 indépendant avec salariés en IVa 43 indépendant sans salariés en IVb
VI	61, 82	0291 (Agent technique des eaux et forêts) en V
VIIa	63, 67, 68, 65, 66	8678 (Femme de ménage dans les collectivités et entreprises) : - en IIIb par défaut (cas le plus fréquent ; le secteur n'est pas disponible pour les chômeurs avant 1975) - en VIIa pour les secteurs suivants : 2820 (Santé humaine) 2821 (Santé animale), 2833 (Hôtels, cafés, restaurants), 3100 (État), 3200 (Armée), 3300 (Coll. loc), 3400 (Sécurité sociale), 3510 (Enseignement privé), 3600 (Ménages), 3700 (Hors économie). 65 qualifiés en VI, contremaîtres en V. 0174 (Pêcheur/Conchyliculteur) et 0471 (Marin pêcheur) en VIIIb
VIIb	10	-

Note : pour cette série, on ne dispose pas des informations nécessaires pour placer les gros indépendants dans la classe I.

Tableau A.2 : Codage de la classe EGP, 1975-1981 (Code des métiers de 1975)

Classe EGP	CSP	Exceptions
I	30, 32, 33, 34	-
II	41, 42, 80, 81, 44	9108 (Journaliste) en I 8302, 8304, 8305, 8314, 8316 (Acheteur, Chef des ventes-de rayon, Pompiste-Vendeur-Autres salariés du commerce, Agent commercial-Représentant de commerce, Voyageur de commerce-Démarcheur-Commis) en IIIa
IIIa	51	8113, 8515, 8623 (Standardiste-Télégraphiste, Gardien, Agent de service) en IIIb
IIIb	53, 70, 71, 72	4615 (Boucher) 8801 (Aide-infirmier) en IIIa 6501, 6514, 6516 (Chauffeurs) en VIIa 8701, 8702, 8703 (Manucure, Coiffeur, Maquilleuse-esthéticienne) en VI 72 indépendant avec salariés en IVa 72 indépendant sans salariés en IVb
IVa	21, 22, 26, 27	sans salariés en IVb
IVb	21, 22, 26, 27	avec salariés en IVa
IVc	00, 23	-
V	43, 60	7502, 7503, 7514, 7515 (Dessinateurs), 7610, 7611, 7614, 7616, 7621, 8891 (Analyste, Programmeur, Arpenteur/Topographe, Géomètre, Expert, Préparateur en pharmacie/Herboriste) en II 43 indépendant avec salariés en IVa 43 indépendant sans salariés en IVb
VI	61, 82	0202 (Agent technique des eaux et forêts) en V
VIIa	63, 67, 68, 65, 66	8622 (Femme de ménage, balayeur, nettoyeur, sans autre indication) - en IIIb par défaut (cas le plus fréquent) - en VIIa pour les secteurs suivants : 2820 (Santé humaine) 2821 (Santé animale), 2833 (Hôtels, cafés, restaurants), 3100 (État), 3200 (Armée), 3300 (Coll. loc.), 3400 (Sécurité sociale), 3510 (Enseignement privé), 3600 (Ménages), 3700 (Hors économie). 65 qualifiés en VI, les contremaîtres en V. 0414 (Pêcheur en eau salée) en VIIIb
VIIIb	10	-

Note : pour cette série, on ne dispose pas des informations nécessaires pour placer les gros indépendants dans la classe I. De même, seule la CSP à un chiffre de l'ancien emploi est connue pour les chômeurs et inactifs. On a donc imputé la catégorie à deux chiffres la plus nombreuse (ou la plus centrale lorsque la catégorie modale était trop spécifique) : 1 → 10 ; 2 → 27 ; 3 → 34 ; 4 → 44 ; 5 → 51 ; 6 → 63 ; 7 → 72 ; 8 → 82.

Tableau A.3 : Codage de la classe EGP, 1982-2002 (Code des PCS 1982)

Classe EGP	PCS	Exceptions
I	31, 33, 34, 35, 37, 38	3531, 3532, 3533, 3534, 3535 en II
II	42, 43, 44, 45, 46	4227, 4231, 4232, 4233 dont la qualification est de niveau ouvrier ou employé en IIIa 4633, 4636 en V 4622, 4623, 4624, 4625, 4627, 4629 en IIIa 46 indépendants avec salariés en IVa 46 indépendants sans salariés en IVb
IIIa	52, 54	5216, 5217, 5222, 5223, 5415, 5417 en IIIb
IIIb	55, 56	5621, 5622 en VI
IVa	23	2310, 2320 en I
IVb	21, 22	avec salariés en IVa
IVc	10, 11, 12, 13	-
V	47, 48	4701, 4711, 4712, 4721, 4722, 4731, 4732, 4733, 4791, 4792, 4793 en II 47 indépendants avec salariés : IVa 47 indépendants sans salariés : IVb
VI	53, 62, 63	5313, 5316 en V 5317 en IIIb
VIIa	64, 65, 67, 68	6511, 6512, 6521, 6522 en VI
VIIb	69	-

Note : en 1990, les noms des variables changent, et on connaît le statut du dernier emploi occupé pour les inactifs et les chômeurs.

Tableau A.4 : Codage de la classe EGP, 2003-2011 (Code des PCS 2003)

Classe EGP	PCS	Exceptions
I	31, 33, 34, 35, 37, 38	354a, 354b, 354c, 354d, 354g en II
II	42, 43, 44, 45, 46	422d, 422e, 423a, 423b, 424a, 425a dont la qualification est de niveau ouvrier ou employé en IIIa 465b, 465c en V 462b, 462d, 462e, 463a, 463b, 463c, 463e en IIIa 46 indépendants avec salariés : IVa 46 indépendants sans salariés : IVb
IIIa	52, 54	525a, 525b, 525c, 525d, 526e, 541d en IIIb
IIIb	55, 56	562a, 562b en VI
IVa	23	231a, 232a en I
IVb	21, 22	avec salariés en IVa
IVc	10, 11, 12, 13	-
V	47, 48	471a, 472a, 472b, 473a, 474a, 478a, 479a en II 47 indépendants avec salariés : IVa 47 indépendants sans salariés : IVb
VI	53, 62, 63	533c, 534a, 534b en V 5317 en IIIb
VIIa	64, 65, 67, 68	651a, 651b, 654a, 655a en VI
VIIb	69	-

Tableau A.5 : Variables et modalités utilisées pour coder la classe EGP des conjoints

		1969-1974	1974-1981	1982-1990	1990-2002	2003-2011
Actif		F 1, 2	FI 1, 2	FI 1 à 3	FI 1, 2	*
CSP/PCS	Actuelle	CSE	CSE	CSE	CSE	CSE
	Ancienne	CSA	CSA	CSA	CSA	CSA
Profession	Actuelle**	P	P	P	P	P
Indépendant	Actuel	ST 2 à 4	ST 1, 2, 3	STATUT 11, 12	STATUT 11, 12	STATUT 11, 12
	Ancien	-	-	-	STANT 1	ATRAV 1
Aide familial	Actuel	ST 1	AIDFAM 1	STATUT 13	STATUT 13	STATUT 13
	Ancien	-	-	-	STANT 1	ATRAV 2
Qualification	Actuelle	Ouvrier qualifié : Q 3 Contremaître : Q 4	Ouvrier qualifié : Q 3 Contremaître : Q 4	Ouvrier- employé : Q 0 à 4	Ouvrier- employé : Q 1, 2, 9	Ouvrier- employé : QPRC 1, 2, 4
A des salariés	Actuel	NSE 1, 3, 6, 8	NSE 1, 2, 3	NSE 1 à 4	NSE 2 à 4	NBSALB 2 à 9
	Ancien	-	-	-	NSEA 2, 3	ANBSAL 2 à 9***
Secteur	Actuel	BCND	BCND			
	Ancien	BCNDA	BCNDA			

* La profession détaillée n'est disponible que pour les actifs occupés.

** Suite à la suppression en 2003 de la question (« filtre initial ») concernant la définition spontanée de son statut par l'interrogé (Goux et Chardon, 2003), il n'est pas possible de conserver la définition extensive de l'activité utilisée avant cette date (plus large que la définition du BIT), qui est la plus appropriée ici étant donné que la quantité d'informations disponibles concernant le dernier emploi occupé est conditionnée par cette variable dans les enquêtes antérieures à 2003. Nous définissons donc après 2003 un actif de la manière la plus large possible, afin de se rapprocher des séries précédentes, comme un individu ayant un statut d'actif la semaine d'enquête (SP00 1 à 4, ou 1 à 3 après 2007), ou remplissant au moins l'une des conditions suivantes : actif au sens du BIT selon l'ancienne définition (ACTANC 1 ou 2) ; ayant recherché un emploi au cours du dernier mois (MREC 1) ; inscrit à Pôle emploi (OFFIC 1 ou 2).

*** Contrairement aux autres séries, ANBSAL renseigne aussi le nombre de salariés de l'entreprise des aides familiaux au chômage ou inactifs au moment de l'enquête.

Un tiret indique l'indisponibilité de l'information, une cellule vide le fait qu'elle n'est pas utilisée pour le recodage.

Tableau A.6 : Variables et modalités utilisées pour coder la classe EGP d'origine

	1982-1990	1990-2002	2003-2011
PCS	CSPP	CSPP	CSPP
Indépendant	STPP 1	STPP 1	OTRAVP 1
Aide familial	STPP 1	STPP 1	OTRAVP 3
A des salariés*	-	NSEPP 2, 3	OSALPP 1 à 9

* Variable non utilisée du fait de son apparition tardive, mais potentiellement disponible pour des études portant sur la période récente.

Un tiret indique l'indisponibilité de l'information.

Dans la série 2003-2011, environ 2 % des pères sont déclarés comme retraités : on ne dispose que du premier chiffre de leur PCS. On a donc imputé la catégorie à deux chiffres (ou la plus centrale lorsque la catégorie modale était trop spécifique) : 71 → 10 ; 72 → 21 ; 74 → 31 ; 75 → 46 ; 77 → 54 ; 78 → 67.

ANNEXE B

LE COEFFICIENT D'ASSOCIATION INTRINSÈQUE ET L'*ODDS RATIO* ABSOLU MOYEN (MAOR), DEUX MESURES DE L'ASSOCIATION INDÉPENDANTES DES MARGES

Il est surprenant qu'en dépit de l'existence d'un corpus important et ancien de littérature concernant les *odds ratios* et les modèles log-linéaires, aucun indice général de l'association existant entre deux variables qualitatives non ordonnées fondée sur cette approche ne se soit imposé. Plusieurs indices fondés sur les contingences de Pearson (tels que le coefficient de contingence carrée moyenne ϕ^2 ou le V de Cramér) sont d'usage courant, mais aucun d'entre eux ne présente la propriété d'indépendance des marges de l'*odds ratio* (Bishop, Fienberg et Holland, [1975] 2007, chap. 11 ; Liebetrau, 1983). Les indices indépendants des marges ayant connu un large succès ne considèrent que le cas particulier des tables à deux lignes et deux colonnes, sous la forme de l'*odds ratio* lui-même, ou encore du coefficient de colligation Y et du coefficient Q de Yule.

Plusieurs indices fondés sur l'*odds ratio* ont été proposés pour mesurer l'intensité d'une association spécifique, d'un intérêt particulier pour l'étude de la mobilité sociale et de l'homogamie : celle appartenant à la diagonale de tables carrées (possédant autant de lignes que de colonnes), correspondant à la tendance à l'hérédité sociale ou à l'endogamie, indépendamment de l'effet des contraintes structurelles. Michel Forsé et Louis Chauvel (1995) ont élaboré l'indice de diagonalité nette, défini comme la moyenne géométrique de tous les *odds ratios* qui peuvent être formés à partir de deux cellules de la diagonale et de deux cellules hors diagonale. Gunn Elisabeth Birkelund et Johan Heldal (2003) ont de leur côté fondé leur *H index* (pour indice d'homogamie) sur les coefficients d'interaction du modèle log-linéaire saturé, contrastant ceux de la diagonale avec ceux du reste de la table.

Ces mesures sont d'une grande utilité pour l'objectif qu'elles se sont fixé ; mais elles ne remplacent pas une mesure plus générale de la force de l'association, qui n'attribuerait aucune attention particulière à la diagonale de la table, et qui s'appliquerait à n'importe quelle table rectangulaire.

Un tel indice a été décrit par Leo Goodman (1996) dans un article majeur réconciliant dans un cadre mathématique commun deux traditions opposées : celle des contingences (ou du χ^2) de Pearson et de l'analyse des correspondances d'un côté ; celle des *odds ratios* et des modèles log-linéaires, log-multiplicatifs et d'association de l'autre. Ce cadre a permis à l'auteur de proposer l'équivalent, dans la seconde tradition, du coefficient de contingence carrée moyenne de Pearson et du V de Cramér : le coefficient d'association intrinsèque. Mais cet indice demeure peu connu et utilisé, à de rares exceptions près (Hout, Brooks et Manza, 1995 ; Breen *et al.*, 2009). Une des raisons de ce manque de succès pourrait tenir à sa difficile interprétation, liée à l'absence de signification simple de l'indice en termes par exemple de chances relatives. Pour tenter de remédier à cette difficulté, nous proposons ici un indice très directement lié au coefficient d'association intrinsèque, mais offrant une interprétation plus simple, dans une échelle familière. L'*odds ratio* absolu moyen (MAOR, pour *Mean Absolute Odds Ratio*) est en effet défini comme l'écart-type géométrique de l'ensemble des *odds ratios* qui peuvent être calculés dans une table. Comme le coefficient d'association intrinsèque, cet indice présente des propriétés très intéressantes pour l'analyse des tables de contingence.

Nous commençons par présenter le cadre général établi par Goodman unifiant la logique pearsonienne et la logique de l'*odds ratio*, pour en dériver le coefficient d'association intrinsèque, indice de l'intensité générale de l'association indépendant des marges de la table. Nous élaborons ensuite l'*odds ratio* absolu moyen (MAOR) dans une version avec pondération uniforme, puis avec pondération arbitraire. Finalement, nous montrons que ces deux types d'indices dérivés de l'*odds ratio* peuvent être utilisés conjointement à des modèles log-multiplicatifs classiques : le modèle UNIDIFF et les modèle d'association RC(M) et RC(M)-L.

1 Un cadre général réunissant tradition pearsonienne et *odds ratios*

Dans cette section, nous présentons le cadre général établi par Goodman (1996), en reprenant la terminologie et les notations de l'article. Ce dernier systématise des résultats développés progressivement par l'auteur une série de travaux (dont notamment Goodman, 1986, 1991). Après avoir défini ce cadre, nous montrons rapidement comment il permet de resituer la tradition issue de Pearson, avant de développer l'élaboration du coefficient d'association intrinsèque comme mesure générale de l'association présente dans une table de contingence suivant la tradition des *odds ratios*.

1.1 Définitions préliminaires

Soient P_{ij} les proportions observées dans la cellule à l'intersection de la ligne i et de la colonne j d'une table de dimensions $I \times J$. Alors les proportions marginales en lignes et en colonnes sont respectivement :

$$P_{i+} = \sum_{j=1}^J P_{ij} \quad \text{et} \quad P_{+j} = \sum_{i=1}^I P_{ij} \quad (1)$$

Suivant cette notation, définissons le ratio de Pearson (dans les termes de Goodman), aussi dénommé ratio d'indépendance, de mobilité ou d'homogamie dans la littérature sociologique :

$$\psi_{ij} = \frac{P_{ij}}{P_{i+}P_{+j}} \quad (2)$$

et une quantité dérivée :

$$R_{ij} = R[\psi_{ij}] \quad (3)$$

avec $R[x]$ une fonction monotone croissante appelée lien d'interaction.

Nous pouvons alors définir l'interaction non pondérée correspondant à une cellule donnée comme :

$$\lambda_{ij} = R_{ij} - R_{i+} - R_{+j} + R_{++} \quad (4)$$

où R_{i+} et R_{+j} sont respectivement les moyennes (non pondérées) en lignes et en colonnes des quantités R_{ij} , et R_{++} est la moyenne de tous les R_{ij} de la table :

$$R_{i+} = \frac{1}{J} \sum_{j=1}^J R_{ij}, \quad R_{+j} = \frac{1}{I} \sum_{i=1}^I R_{ij}, \quad \text{et} \quad R_{++} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J R_{ij} \quad (5)$$

Nous pouvons définir de manière similaire l'interaction pondérée comme :

$$\tilde{\lambda}_{ij} = R_{ij} - \tilde{R}_{i+} - \tilde{R}_{+j} + \tilde{R}_{++} \quad (6)$$

où \tilde{R}_{i+} et \tilde{R}_{+j} sont respectivement les moyennes pondérées par les marges des quantités R_{ij} , et \tilde{R}_{++} est la moyenne pondérée de tous les R_{ij} de la table. Ces termes sont définis comme :

$$\tilde{R}_{i+} = \sum_{j=1}^J R_{ij} P_{+j}, \quad \tilde{R}_{+j} = \sum_{i=1}^I R_{ij} P_{i+}, \quad \text{et} \quad \tilde{R}_{++} = \sum_{i=1}^I \sum_{j=1}^J R_{ij} P_{i+} P_{+j} \quad (7)$$

Notons que si nous retenons ici par simplicité des pondérations égales aux proportions marginales de la table, tout jeu de poids strictement positifs dont la somme est égale à 1 peut être utilisé. C'est le cas notamment des pondérations uniformes, égales respectivement à $1/I$ pour les lignes et $1/J$ pour les colonnes, fréquemment utilisées, comme nous le verrons plus bas.

D'après les équations (4) et (5), nous pouvons voir que les sommes en lignes et en colonnes des λ_{ij} sont nulles :

$$\sum_{i=1}^I \lambda_{ij} = 0 \quad \text{et} \quad \sum_{j=1}^J \lambda_{ij} = 0 \quad (8)$$

et d'après les équations (6) et (7), nous pouvons voir que c'est aussi le cas des sommes pondérées en lignes et en colonnes des $\tilde{\lambda}_{ij}$:

$$\sum_{i=1}^I \tilde{\lambda}_{ij} P_{i+} = 0 \quad \text{avec} \quad j = 1, \dots, J \quad \text{et} \quad \sum_{j=1}^J \tilde{\lambda}_{ij} P_{+j} = 0 \quad \text{avec} \quad i = 1, \dots, I \quad (9)$$

Afin d'obtenir une mesure générale de l'intensité de l'association au sein de la table, Goodman (1996, sect. 7) a proposé un indice généralisé de non-indépendance. Dans sa version non pondérée, notée λ , il est égal à la norme euclidienne des λ_{ij} :

$$\lambda = \sqrt{\sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2} \quad (10)$$

Et dans sa version pondérée, notée $\tilde{\lambda}$, à l'écart-type pondéré des $\tilde{\lambda}_{ij}$ ¹ :

$$\tilde{\lambda} = \sqrt{\sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+} P_{+j}} \quad (11)$$

À la place d'une pondération marginale, la pondération uniforme déjà décrite donne une troisième version communément retenue du coefficient, notée λ^\dagger , dont on peut montrer d'après (5) et (7) ainsi que (8) et (9) qu'elle correspond à l'écart-type des λ_{ij} :

$$\lambda^\dagger = \sqrt{\frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2} = \frac{\lambda}{\sqrt{IJ}} \quad (12)$$

Les définitions (10), (11) et (12) ci-dessus font immédiatement apparaître une propriété intéressante des différentes versions de l'indice d'écart à l'indépendance : le carré de cet indice, défini comme une somme de contributions par cellule, peut aussi se décomposer en contributions par ligne et par colonne, en prenant simplement les sommes des contributions par cellule correspondantes. Soit respectivement, pour les versions non pondérée, avec pondération par les marges, et avec pondération uniforme :

$$\sum_{j=1}^J \lambda_{ij}^2 \text{ et } \sum_{i=1}^I \lambda_{ij}^2, \quad \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{+j} \text{ et } \sum_{i=1}^I \tilde{\lambda}_{ij}^2 P_{i+}, \quad \frac{1}{J} \sum_{j=1}^J \lambda_{ij}^2 \text{ et } \frac{1}{I} \sum_{i=1}^I \lambda_{ij}^2 \quad (13)$$

Nous commençons par montrer que la version pondérée de l'indice, notée $\tilde{\lambda}$, constitue une généralisation du coefficient de contingence carrée moyenne φ^2 de Pearson, avant de

1 La somme des poids étant égale à 1, la version pondérée correspond à une moyenne, alors la version non pondérée équivaut à une somme.

développer plus longuement comment une version dérivée des *odds ratios* peut être établie dans ce cadre².

1.2 Le coefficient de contingence carrée moyenne de Pearson

Si nous définissons R comme étant la fonction identité, soit $R[x] = x$, et par conséquent :

$$R_{ij} = \psi_{ij} = \frac{P_{ij}}{P_{i+}P_{+j}} \quad (14)$$

alors l'équation (7) implique :

$$\tilde{R}_{i+} = \tilde{R}_{+j} = \tilde{R}_{++} = 1 \quad (15)$$

et donc :

$$\tilde{\lambda}_{ij} = R_{ij} - 1 - 1 + 1 = \psi_{ij} - 1 = \frac{P_{ij} - P_{i+}P_{+j}}{P_{i+}P_{+j}} \quad (16)$$

Dans ce cas, nous trouvons que les $\tilde{\lambda}_{ij}$ sont, dans le vocabulaire de Goodman, les contingences de Pearson, ou encore les racines carrées des résidus de Pearson. L'indice général d'écart à l'indépendance défini à l'équation (11) comme $\tilde{\lambda}$ apparaît donc égal à la racine carrée du coefficient de contingence carrée moyenne de Pearson (ou φ^2) :

$$\tilde{\lambda}^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(P_{ij} - P_{i+}P_{+j})^2}{P_{i+}P_{+j}} = \varphi^2 = \frac{\chi^2}{N} \quad (17)$$

avec N l'effectif total de la table.

² Nous laisserons ici de côté la démonstration du lien avec les coefficients Y et Q de Yule qui, moins communément utilisés, présentent des propriétés similaires à celles de l'*odds ratio*.

1.3 Le coefficient d'association intrinsèque

Le cadre général établi par Goodman est particulièrement fécond du fait qu'il autorise l'extension des mesures développées dans une démarche issue de Pearson, dans laquelle le lien d'interaction R est la fonction identité, à la tradition des *odds ratios* et de la modélisation log-linéaire. L'auteur montre en effet que cette seconde approche peut être obtenue en définissant R comme étant le logarithme naturel, soit $R[x] = \log x$. Dans ce nouveau cadre, on a :

$$R_{ij} = \log \psi_{ij} = \log P_{ij} - \log P_{i+} - \log P_{+j} \quad (18)$$

Les équations (5) (p. 528) et (18) impliquent que :

$$\begin{aligned} R_{i+} &= \frac{1}{J} \sum_{j=1}^J \log P_{ij} - \frac{1}{J} \sum_{j=1}^J \log P_{+j} - \log P_{i+} \\ R_{+j} &= \frac{1}{I} \sum_{i=1}^I \log P_{ij} - \frac{1}{I} \sum_{i=1}^I \log P_{i+} - \log P_{+j} \\ R_{++} &= \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log P_{ij} - \frac{1}{I} \sum_{i=1}^I \log P_{i+} - \frac{1}{J} \sum_{j=1}^J \log P_{+j} \end{aligned} \quad (19)$$

Par conséquent, en combinant les équations (4) (p. 527), (18) et (19), on obtient :

$$\begin{aligned} \lambda_{ij} &= \left(\log P_{ij} - \log P_{i+} - \log P_{+j} \right) \\ &\quad - \left(\frac{1}{J} \sum_{j=1}^J \log P_{ij} - \frac{1}{J} \sum_{j=1}^J \log P_{+j} - \log P_{i+} \right) \\ &\quad - \left(\frac{1}{I} \sum_{i=1}^I \log P_{ij} - \frac{1}{I} \sum_{i=1}^I \log P_{i+} - \log P_{+j} \right) \\ &\quad + \left(\frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log P_{ij} - \frac{1}{I} \sum_{i=1}^I \log P_{i+} - \frac{1}{J} \sum_{j=1}^J \log P_{+j} \right) \\ &= \log P_{ij} - \frac{1}{J} \sum_{j=1}^J \log P_{ij} - \frac{1}{I} \sum_{i=1}^I \log P_{ij} + \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log P_{ij} \end{aligned} \quad (20)$$

Les λ_{ij} sont donc égaux aux coefficients d'interaction du modèle log-linéaire saturé (justifiant ainsi la notation choisie)³, qui sont plus souvent présentés sous la forme (Bishop, Fienberg et Holland, [1975] 2007, chap. 2 ; Agresti, [1990] 2002, chap. 5) :

$$\log P_{ij} = \lambda^0 + \lambda_i^I + \lambda_j^J + \lambda_{ij} \quad (21)$$

avec

$$\lambda^0 = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log P_{ij}, \quad \lambda_i^I = \frac{1}{J} \sum_{j=1}^J \log P_{ij} - \lambda^0, \quad \lambda_j^J = \frac{1}{I} \sum_{i=1}^I \log P_{ij} - \lambda^0 \quad (22)$$

$$\sum_{i=1}^I \lambda_i^I = \sum_{j=1}^J \lambda_j^J = 0, \text{ et comme vu plus haut } \sum_{i=1}^I \lambda_{ij} = \sum_{j=1}^J \lambda_{ij} = 0$$

Ainsi, l'*odds ratio* contrastant les lignes i et i' et les colonnes j et j' est égal à :

$$\theta_{ij,i'j'} = \exp \left[\lambda_{ij} + \lambda_{i'j'} - \lambda_{i'j} - \lambda_{ij'} \right] \quad (23)$$

Comme ce dernier, les λ_{ij} présentent la propriété d'indépendance des marges qui est au fondement de la modélisation log-linéaire : ils ne sont pas affectés par la multiplication des lignes et des colonnes par des valeurs arbitraires. Ainsi, le remplacement de P_{ij} par $\delta_i \gamma_j P_{ij}$ dans l'équation (20) ne change pas l'expression des λ_{ij} :

3 Notons qu'une décomposition de la variance des effectifs prédits par un modèle log-linéaire en composantes correspondant à chacun de ses paramètres avait déjà été réalisée par Robert Erikson et John Goldthorpe (1992, p. 228), quoique dans une optique assez différente.

$$\begin{aligned}
\lambda'_{ij} &= \log \delta_i \gamma_j P_{ij} - \frac{1}{J} \sum_{j=1}^J \log \delta_i \gamma_j P_{ij} - \frac{1}{I} \sum_{i=1}^I \log \delta_i \gamma_j P_{ij} + \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log \delta_i \gamma_j P_{ij} \\
&= (\log \delta_i + \log \gamma_j + \log P_{ij}) \\
&\quad - \left(\log \delta_i + \frac{1}{J} \sum_{j=1}^J \log \gamma_j + \frac{1}{J} \sum_{j=1}^J \log P_{ij} \right) \\
&\quad - \left(\log \gamma_j + \frac{1}{I} \sum_{i=1}^I \log \delta_i + \frac{1}{I} \sum_{i=1}^I \log P_{ij} \right) \\
&\quad + \left(\frac{1}{I} \sum_{i=1}^I \log \delta_i + \frac{1}{J} \sum_{j=1}^J \log \gamma_j + \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log P_{ij} \right) \\
&= \log P_{ij} - \frac{1}{J} \sum_{j=1}^J \log P_{ij} - \frac{1}{I} \sum_{i=1}^I \log P_{ij} + \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log P_{ij} \\
&= \lambda_{ij}
\end{aligned} \tag{24}$$

Des propriétés équivalentes sont obtenues concernant l'interaction pondérée. En effet, les équations (7) (p. 528) et (18) impliquent que :

$$\begin{aligned}
\tilde{R}_{i+} &= \sum_{j=1}^J P_{+j} \log P_{ij} - \sum_{j=1}^J P_{+j} \log P_{+j} - \log P_{i+} \\
\tilde{R}_{+j} &= \sum_{i=1}^I P_{i+} \log P_{ij} - \sum_{i=1}^I P_{i+} \log P_{i+} - \log P_{+j} \\
\tilde{R}_{++} &= \sum_{i=1}^I \sum_{j=1}^J P_{i+} P_{+j} \log P_{ij} - \sum_{i=1}^I P_{i+} \log P_{i+} - \sum_{j=1}^J P_{+j} \log P_{+j}
\end{aligned} \tag{25}$$

Par conséquent, en combinant les équations (6) (p. 528), (18) et (25) on trouve :

$$\tilde{\lambda}_{ij} = \log P_{ij} - \sum_{j=1}^J P_{+j} \log P_{ij} - \sum_{i=1}^I P_{i+} \log P_{ij} + \sum_{i=1}^I \sum_{j=1}^J P_{i+} P_{+j} \log P_{ij} \tag{26}$$

On obtient ainsi une version originale, avec pondération marginale, du modèle log-linéaire saturé :

$$\log P_{ij} = \tilde{\lambda}^0 + \tilde{\lambda}_i^I + \tilde{\lambda}_j^J + \tilde{\lambda}_{ij} \tag{27}$$

avec

$$\begin{aligned} \tilde{\lambda}^0 &= \sum_{i=1}^I \sum_{j=1}^J P_{i+} P_{+j} \log P_{ij}, \quad \tilde{\lambda}_i^I = \sum_{j=1}^J P_{+j} \log P_{ij} - \tilde{\lambda}^0, \quad \tilde{\lambda}_j^J = \sum_{i=1}^I P_{i+} \log P_{ij} - \tilde{\lambda}^0 \\ \sum_{i=1}^I \tilde{\lambda}_i^I P_{i+} &= \sum_{j=1}^J \tilde{\lambda}_j^J P_{+j} = 0, \text{ et comme vu plus haut } \sum_{i=1}^I P_{i+} \tilde{\lambda}_{ij} = \sum_{j=1}^J P_{+j} \tilde{\lambda}_{ij} = 0 \end{aligned} \quad (28)$$

Comme dans le modèle log-linéaire saturé classique, l'*odds ratio* contrastant les lignes i et i' et les colonnes j et j' est égal à :

$$\theta_{ij,i'j'} = \exp \left[\tilde{\lambda}_{ij} + \tilde{\lambda}_{i'j'} - \tilde{\lambda}_{i'j} - \tilde{\lambda}_{ij'} \right] \quad (29)$$

Cependant, contrairement à l'*odds ratio* et à leur équivalent non pondéré, les coefficients d'interaction pondérés $\tilde{\lambda}_{ij}$ ne sont pas eux-mêmes indépendants des marges lorsqu'une pondération marginale est utilisée. En revanche, les $\tilde{\lambda}_{ij}$ sont indépendants des marges lorsque la pondération retenue l'est aussi ; c'est notamment le cas de la pondération uniforme définie plus haut, et que l'on retrouve par l'utilisation de poids appropriés.

Concluons cette section par des remarques concernant les indices d'intensité générale de l'association qui peuvent être dérivés de ces deux versions pondérée et non pondérée. Dans la logique de cette section, qui est celle des *odds ratios*, l'indice d'intensité de l'association défini en (10) (p. 529) dans sa version non pondérée possède la propriété d'indépendance des marges de la table, étant donné qu'il ne dépend que des λ_{ij} . En revanche, il ne la possède à l'évidence pas dans sa version pondérée par les marges définie en (11), les $\tilde{\lambda}_{ij}$ n'étant pas indépendants des marges. Néanmoins, si des poids arbitraires indépendants des marges sont utilisés, la version pondérée de l'indice est elle aussi indépendante des marges. C'est notamment le cas lorsque des pondérations uniformes sont utilisées, comme dans l'équation (12).

Pour des raisons qui apparaîtront clairement plus bas, nous désignons cet indice, issu de l'approche des modèles log-linéaires, sous le nom de coefficient d'association intrinsèque. Notons cependant que son extension à la régression logistique a été présentée sous le nom d'indice κ (Hout, Brooks et Manza, 1995 ; Breen *et al.*, 2009). Ce coefficient est exprimé sur l'échelle du logarithme des *odds ratios* : il est égal à 0 à l'indépendance (tous les λ_{ij} sont nuls, les *odds ratios* sont égaux à 1), et ne possède pas de borne supérieure (à l'instar des λ_{ij} et des *odds ratios*). Ainsi, si ce coefficient d'association intrinsèque présente la propriété recherchée d'indépendance des marges de la table, et constitue donc un outil important, il

n'offre aucune interprétation substantive aisée, autre qu'en termes d'écart-type des coefficients log-linéaires d'interaction. C'est pourquoi il semble intéressant de se ramener à l'échelle plus familière, et liée à des rapports entre proportions, des *odds ratios*.

2 L'*odds ratio* absolu moyen (MAOR)

Dans cette section, nous montrons qu'un indice à l'interprétation plus aisée que le coefficient d'association intrinsèque, que nous dénommerons *odds ratio* absolu moyen (MAOR, pour *Mean Absolute Odds Ratio*), est directement relié à ce dernier, et en partage les propriétés intéressantes, tout en présentant l'avantage d'un lien direct avec l'*odds ratio*. Le principe en est simple. L'*odds ratio* constitue une mesure de l'intensité de l'association indépendante des marges adéquate dans le cas d'une table de dimensions 2×2 , ou dans l'optique de comparer deux catégories en lignes et deux catégories en colonnes. Un de ses intérêts est d'offrir une interprétation en termes de chances relatives directement en lien avec les proportions observées. Nous nous proposons ici de généraliser cette quantité de manière à mesurer l'intensité générale de l'association dans une table de dimensions $I \times J$.

Pour cela, nous définissons le MAOR comme l'écart-type⁴ géométrique de l'ensemble des *odds ratios* calculables à partir d'une table. De même que le coefficient d'association intrinsèque est égal à l'écart-type des coefficients d'interaction λ_{ij} ou $\tilde{\lambda}_{ij}$, il s'agit ici de mesurer l'écart de l'*odds ratio* par rapport à la valeur de référence (ici, 1), mais sur une échelle multiplicative plutôt que linéaire. En effet, l'écart-type géométrique est l'équivalent de l'écart-type arithmétique classique sur une échelle multiplicative, c'est-à-dire à l'exponentielle d'un écart-type fondé sur les logarithmes des *odds ratios*.

L'une des manières d'énumérer l'ensemble des *odds ratios* pouvant être construits à partir d'une table de dimensions $I \times J$ est de prendre pour référence successivement chaque cellule de la table, et de calculer pour chacune l'ensemble des $(I-1)(J-1)$ *odds ratios* impliquant cette cellule et toutes les cellules situées sur une autre ligne et une autre colonne (*spanning cell odds ratios*, cf. Rudas, 1998)⁵. Cette approche revient à construire

4 Une autre approche possible serait de recourir à l'écart moyen absolu (*mean absolute deviation*, *MAD*), qui retient la valeur absolue (norme 1) plutôt que le carré (norme 2). L'utilisation de l'écart-type permet de se rattacher au cadre élaboré notamment par Goodman dans l'article déjà cité (1996), mais aussi par l'ensemble de la littérature portant sur les modèles d'association.

5 Il serait strictement équivalent, mais moins efficace en termes de redondances, de choisir arbitrairement un *odds ratio* local (voire de les retenir tous) et de calculer sa valeur pour chacune des $I!J!$ permutations

$IJ(I-1)(J-1)$ odds ratios, dont la plupart sont redondants puisque le *basic set* (Goodman, 1969 ; Rudas, 1998) des $(I-1)(J-1)$ *spanning cell odds ratios* correspondant à n'importe quelle cellule suffit à recalculer tous les autres. En outre, cette série d'*odds ratios* contient exactement 4 fois chacun des $IJ(I-1)(J-1)/4$ log-odds ratios au carré différents qui peuvent être construits à partir de la table. En effet, en notant $\theta_{ij,i'j'}$ l'*odds ratio* contrastant les cellules à l'intersection des lignes i et i' et des colonnes j et j' , on a :

$$\log \theta_{ij,i'j'} = \log \frac{P_{ij}P_{i'j'}}{P_{i'j}P_{ij}} = \lambda_{ij} + \lambda_{i'j'} - \lambda_{i'j} - \lambda_{ij'} \quad (30)$$

et donc :

$$(\log \theta_{ij,i'j'})^2 = (\log \theta_{i'j',ij})^2 = (\log \theta_{i'j,ij'})^2 = (\log \theta_{ij',i'j})^2 \quad (31)$$

En suivant la démarche décrite, on retiendra donc 4 fois le même jeu de 4 cellules, ce qui revient à compter 4 fois chaque log-odds ratio carré. Ces redondances ne sont pas un problème, puisque l'écart-type n'est pas affecté par la répétition de l'ensemble des valeurs un nombre égal de fois. En outre, il s'agit uniquement ici de proposer une définition substantive du MAOR, dont nous allons maintenant montrer qu'elle rejoint une expression bien plus simple en termes de calcul.

2.1 MAOR avec pondération uniforme

Ainsi, nous pouvons définir une première version avec pondération uniforme du MAOR⁶, avant de la généraliser dans la section suivante à des poids arbitraires. Elle est égale à l'exponentielle de l'écart-type (arithmétique) non pondéré des $IJ(I-1)(J-1)$ log-odds ratios construits en prenant successivement pour référence chacune des cellules de la table :

$$\begin{aligned} \log^2 MAOR &= \frac{1}{IJ(I-1)(J-1)} \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\log \theta_{ij,i'j'})^2 \\ &= \frac{1}{IJ(I-1)(J-1)} \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\lambda_{ij} + \lambda_{i'j'} - \lambda_{i'j} - \lambda_{ij'})^2 \end{aligned} \quad (32)$$

possibles des lignes et des colonnes.

6 Il est aussi possible de définir une version non pondérée, équivalente du coefficient d'association intrinsèque non pondéré présenté plus haut. Néanmoins, cette version, qui n'est pas un écart-type, n'offre pas d'interprétation en termes d'*odds ratios*, et (en tant que produit plutôt que moyenne géométrique) prend des valeurs trop élevées pour être d'une quelconque utilité pratique.

Cette équation peut être développée de la manière suivante :

$$\begin{aligned}
& IJ(I-1)(J-1) \log^2 MAOR \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\lambda_{ij} + \lambda_{i'j'} - \lambda_{i'j} - \lambda_{ij'})^2 \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\lambda_{ij}^2 + \lambda_{i'j'}^2 + \lambda_{i'j}^2 + \lambda_{ij'}^2) \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\lambda_{ij}\lambda_{i'j'} + \lambda_{i'j}\lambda_{ij'} - \lambda_{ij}\lambda_{i'j} - \lambda_{ij}\lambda_{ij'} - \lambda_{i'j'}\lambda_{i'j} - \lambda_{i'j'}\lambda_{ij'}) \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\lambda_{ij}^2 + \lambda_{i'j'}^2 + \lambda_{i'j}^2 + \lambda_{ij'}^2) \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \lambda_{ij}\lambda_{i'j'} \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\lambda_{i'j}\lambda_{ij'} - \lambda_{ij}\lambda_{i'j} - \lambda_{ij}\lambda_{ij'}) \\
&\quad - 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \lambda_{i'j'}\lambda_{i'j} \\
&\quad - 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \lambda_{i'j'}\lambda_{ij'}
\end{aligned} \tag{33}$$

Or, les termes de cette expression peuvent être ramenés à une forme commune :

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\lambda_{ij}^2 + \lambda_{i'j'}^2 + \lambda_{i'j}^2 + \lambda_{ij'}^2) = 4(I-1)(J-1) \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \tag{34}$$

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \lambda_{ij}\lambda_{i'j'} = \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \quad \text{car} \quad \sum_{j' \neq j} \sum_{i' \neq i} \lambda_{i'j'} = \sum_{j' \neq j} -\lambda_{ij'} = \sum_{i' \neq i} -\lambda_{i'j} = \lambda_{ij} \tag{35}$$

$$\begin{aligned}
\sum_{i=1}^I \sum_{j=1}^J \sum_{j' \neq j} \sum_{i' \neq i} (\lambda_{i'j}\lambda_{ij'} - \lambda_{ij}\lambda_{i'j} - \lambda_{ij}\lambda_{ij'}) &= \sum_{i=1}^I \sum_{j=1}^J (\lambda_{ij}^2 + (I-1)\lambda_{ij}^2 + (J-1)\lambda_{ij}^2) \\
&\quad \text{car} \quad \sum_{j' \neq j} \lambda_{ij'} = \sum_{i' \neq i} \lambda_{i'j} = -\lambda_{ij}
\end{aligned} \tag{36}$$

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \lambda_{i'j'} \lambda_{i'j} = -(I-1) \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \quad \text{car} \quad \sum_{j' \neq j} \lambda_{i'j'} \lambda_{i'j} = -\lambda_{i'j}^2 \quad (37)$$

$$\text{et} \quad \sum_{i=1}^I \sum_{i' \neq i} \lambda_{i'j}^2 = (I-1) \sum_{i=1}^I \lambda_{ij}^2$$

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \lambda_{i'j'} \lambda_{ij'} = -(J-1) \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \quad \text{car} \quad \sum_{i' \neq i} \lambda_{i'j'} \lambda_{ij'} = -\lambda_{ij'}^2 \quad (38)$$

$$\text{et} \quad \sum_{j=1}^J \sum_{j' \neq j} \lambda_{ij'}^2 = (J-1) \sum_{j=1}^J \lambda_{ij}^2$$

Ainsi, l'équation (33) se simplifie comme suit :

$$\begin{aligned} IJ(I-1)(J-1) \log^2 MAOR &= 4(I-1)(J-1) \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \\ &+ 2 \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \\ &+ 2(1 + (I-1) + (J-1)) \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \\ &+ 2((I-1) + (J-1)) \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \\ &= 4IJ \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \\ &= 4IJ\lambda^2 \end{aligned} \quad (39)$$

On trouve donc que le MAOR avec pondération uniforme, défini comme l'écart-type géométrique de tous les *odds ratios* de la table, est égal à :

$$MAOR = \exp \sqrt{\frac{4}{(I-1)(J-1)}} \lambda = \exp \sqrt{\frac{4IJ}{(I-1)(J-1)}} \lambda^\dagger \quad (40)$$

Le lien avec le coefficient d'association intrinsèque est donc direct : il ne tient qu'au passage de l'échelle du logarithme des *odds ratios* à celle des *odds ratios*, grâce à la fonction

exponentielle, après multiplication par un facteur qui dépend uniquement des dimensions de la table⁷. Le recours à la fonction exponentielle peut se comprendre par le fait que, contrairement au coefficient de contingence carrée moyenne φ^2 de Pearson, pour lequel la fonction de lien d'interaction $R[x]$ définie à l'équation (3) (p. 527) est l'identité, l'approche par les *odds ratios* repose sur un lien logarithmique. Ainsi, il s'agit ici de revenir à l'échelle naturelle en appliquant la réciproque de la fonction de lien, soit la fonction exponentielle.

En se plaçant dans le cas particulier d'une table carrée de dimensions 2×2 , on peut vérifier le lien direct entre le MAOR avec pondération uniforme et l'*odds ratio*⁸. En effet, en posant, sans perte de généralité⁹, que l'*odds ratio* contrastant les lignes 1 et 2 et les colonnes 1 et 2 est supérieur à l'unité, soit $\theta_{11,22} > 1$ et $\lambda_{11} > 0$, et puisque $\lambda_{11} = \lambda_{22} = -\lambda_{12} = -\lambda_{21}$, on a :

$$\begin{aligned}
 MAOR &= \exp \sqrt{4 \left(\lambda_{11}^2 + \lambda_{12}^2 + \lambda_{21}^2 + \lambda_{22}^2 \right)} \\
 &= \exp \sqrt{4 \times 4\lambda_{11}^2} \\
 &= \exp [4\lambda_{11}] \\
 &= \exp [\lambda_{11} + \lambda_{22} - \lambda_{12} - \lambda_{21}] \\
 &= \exp [\log \theta_{11,22}] \\
 &= \theta_{11,22} = \theta_{22,11} = \frac{1}{\theta_{12,21}} = \frac{1}{\theta_{21,12}}
 \end{aligned} \tag{41}$$

Le MAOR avec pondération uniforme est donc égal à l'*odds ratio* supérieur à 1 de la table, ou à l'inverse de l'*odds ratio* inférieur à 1. Il constitue ainsi une généralisation de l'*odds ratio* aux tables de plus de quatre cellules.

Après avoir défini la version pondérée du MAOR dans la section suivante, nous verrons que cette dernière permet de retrouver les démonstrations que nous venons de réaliser en utilisant des pondérations uniformes ($1/I$ pour les lignes et $1/J$ pour les colonnes).

7 On peut remarquer que la constante 4 correspond au fait que chaque *odds ratio* est calculé à partir de 4 coefficients d'interaction λ_{ij} – i.e. à partir de 4 cellules.

8 Goodman (1996, p. 411) avait déjà souligné que dans le cas d'une telle table, le coefficient d'association intrinsèque avec pondération uniforme était proportionnel au log-*odds ratio*, mais sans en tirer un indice général de l'association.

9 En effet, une permutation des lignes ou des colonnes permet de choisir lequel des *odds ratios* remplira cette condition.

2.2 MAOR avec pondération arbitraire

Suivant la démarche déjà retenue ci-dessus, nous définissons maintenant le MAOR avec pondération arbitraire, qui généralise les résultats de la section précédente. Comme plus haut, par simplicité, nous présentons le cas particulier de la pondération marginale, mais les démonstrations sont valables, sauf mention contraire, pour tout jeu de poids strictement positifs dont la somme est égale à 1, du moment que les coefficients d'interaction $\tilde{\lambda}$ ont été calculés avec les mêmes poids. En effet, l'intérêt de cette version du MAOR n'est pas tant de permettre d'utiliser des pondérations marginales dans une table à deux dimensions – ce qui ferait perdre la propriété d'indépendance des marges –, mais plutôt, comme nous le verrons plus loin, d'utiliser une pondération par la moyenne des marges dans une table à trois dimensions.

Cette seconde version du MAOR, notée \widetilde{MAOR} , est définie comme l'exponentielle de l'écart-type (arithmétique) pondéré de tous les log-odds ratios d'une table de dimensions $I \times J$:

$$\begin{aligned} \log^2 \widetilde{MAOR} &= \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\log \theta_{ij, i'j'})^2 P_{i+} P_{+j} P_{i'+} P_{+j'}}{\sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} P_{i+} P_{+j} P_{i'+} P_{+j'}} \\ &= \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\tilde{\lambda}_{ij} + \tilde{\lambda}_{i'j'} - \tilde{\lambda}_{i'j} - \tilde{\lambda}_{ij'})^2 P_{i+} P_{+j} P_{i'+} P_{+j'}}{\sum_{i=1}^I \sum_{j=1}^J P_{i+} (1 - P_{i+}) P_{+j} (1 - P_{+j})} \end{aligned} \quad (42)$$

En suivant la même démarche que pour l'indice avec pondération uniforme, on trouve l'équivalent de l'équation (33) (p. 537) :

$$\begin{aligned}
& \sum_{i=1}^I \sum_{j=1}^J P_{i+}(1 - P_{i+})P_{+j}(1 - P_{+j}) \log^2 \widetilde{MAOR} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} (\tilde{\lambda}_{ij} + \tilde{\lambda}_{i'j'} - \tilde{\lambda}_{i'j} - \tilde{\lambda}_{ij'})^2 P_{i+}P_{+j}P_{i'+}P_{+j'} \\
&= 4 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}P_{+j}(1 - P_{i+})(1 - P_{+j}) \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{ij} \tilde{\lambda}_{i'j'} P_{i+}P_{+j}P_{i'+}P_{+j'} \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{i'j} \tilde{\lambda}_{ij'} P_{i+}P_{+j}P_{i'+}P_{+j'} \\
&\quad - 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{ij} \tilde{\lambda}_{i'j} P_{i+}P_{+j}P_{i'+}P_{+j'} \\
&\quad - 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{ij} \tilde{\lambda}_{ij'} P_{i+}P_{+j}P_{i'+}P_{+j'} \\
&\quad - 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{i'j'} \tilde{\lambda}_{i'j} P_{i+}P_{+j}P_{i'+}P_{+j'} \\
&\quad - 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{i'j'} \tilde{\lambda}_{ij'} P_{i+}P_{+j}P_{i'+}P_{+j'}
\end{aligned} \tag{43}$$

Les termes de cette expression peuvent être ramenés de la même manière à une forme commune :

$$\begin{aligned}
& \sum_{i=1}^I \sum_{j=1}^J \sum_{j' \neq j} \sum_{i' \neq i} \tilde{\lambda}_{ij} \tilde{\lambda}_{i'j'} P_{i+}P_{+j}P_{i'+}P_{+j'} = \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j}^2 \\
&\quad \text{car } \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{i'j'} P_{i+}P_{+j}P_{i'+}P_{+j'} = - \sum_{i' \neq i} \tilde{\lambda}_{i'j} P_{i+}P_{+j}^2 P_{i'+} = \tilde{\lambda}_{ij} P_{i+}^2 P_{+j}^2
\end{aligned} \tag{44}$$

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{i'j} \tilde{\lambda}_{ij'} P_{i+P+j} P_{i'+P+j'} &= \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j}^2 \\ \text{car } \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{i'j} \tilde{\lambda}_{ij'} P_{i+P+j} P_{i'+P+j'} &= - \sum_{i' \neq i} \tilde{\lambda}_{ij} \tilde{\lambda}_{i'j} P_{i+P+j} P_{i'+P+j'} = \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j}^2 \end{aligned} \quad (45)$$

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{ij} \tilde{\lambda}_{i'j'} P_{i+P+j} P_{i'+P+j'} &= - \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j} (1 - P_{+j}) \\ \text{car } \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{i'j} \tilde{\lambda}_{ij'} P_{i+P+j} P_{i'+P+j'} &= - \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j} \sum_{j' \neq j} P_{+j'} = - \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j} (1 - P_{+j}) \end{aligned} \quad (46)$$

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \sum_{j' \neq j} \sum_{i' \neq i} \tilde{\lambda}_{ij} \tilde{\lambda}_{i'j'} P_{i+P+j} P_{i'+P+j'} &= - \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+} P_{+j}^2 (1 - P_{i+}) \\ \text{car } \sum_{i' \neq i} \sum_{j' \neq j} \tilde{\lambda}_{i'j} \tilde{\lambda}_{ij'} P_{i+P+j} P_{i'+P+j'} &= - \tilde{\lambda}_{ij}^2 P_{i+} P_{+j}^2 \sum_{i' \neq i} P_{i'+} = - \tilde{\lambda}_{ij}^2 P_{i+} P_{+j}^2 (1 - P_{i+}) \end{aligned} \quad (47)$$

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \sum_{j' \neq j} \sum_{i' \neq i} \tilde{\lambda}_{i'j'} \tilde{\lambda}_{ij} P_{i+P+j} P_{i'+P+j'} &= - \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+} P_{+j}^2 (1 - P_{i+}) \\ \text{car } \sum_{j' \neq j} \tilde{\lambda}_{i'j'} \tilde{\lambda}_{ij} P_{i+P+j} P_{i'+P+j'} &= - \tilde{\lambda}_{i'j}^2 P_{i+} P_{+j}^2 P_{i'+} \\ \text{et } \sum_{i=1}^I \sum_{i' \neq i} \tilde{\lambda}_{i'j}^2 P_{i+} P_{+j}^2 P_{i'+} &= \sum_{i=1}^I \tilde{\lambda}_{ij}^2 P_{i+} P_{+j}^2 (1 - P_{i+}) \end{aligned} \quad (48)$$

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \sum_{j' \neq j} \sum_{i' \neq i} \tilde{\lambda}_{i'j'} \tilde{\lambda}_{ij} P_{i+P+j} P_{i'+P+j'} &= - \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+} P_{+j}^2 (1 - P_{+j}) \\ \text{car } \sum_{i' \neq i} \tilde{\lambda}_{i'j'} \tilde{\lambda}_{ij} P_{i+P+j} P_{i'+P+j'} &= - \tilde{\lambda}_{i'j}^2 P_{i+} P_{+j}^2 P_{+j'} \\ \text{et } \sum_{j=1}^J \sum_{j' \neq j} \tilde{\lambda}_{i'j}^2 P_{i+} P_{+j}^2 P_{+j'} &= \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+} P_{+j}^2 (1 - P_{+j}) \end{aligned} \quad (49)$$

Ainsi, l'équation (43) se simplifie comme suit :

$$\begin{aligned}
& \sum_{i=1}^I \sum_{j=1}^J P_{i+}(1 - P_{i+})P_{+j}(1 - P_{+j}) \log^2 \widetilde{MAOR} \\
&= 4 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}P_{+j}(1 - P_{i+})(1 - P_{+j}) \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j}^2 \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j}^2 \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j}(1 - P_{+j}) \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}P_{+j}^2(1 - P_{i+}) \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}P_{+j}^2(1 - P_{i+}) \\
&\quad + 2 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}^2 P_{+j}(1 - P_{+j}) \\
&= 4 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}P_{+j} \left((1 - P_{i+})(1 - P_{+j}) + P_{i+}P_{+j} + P_{+j}(1 - P_{i+}) + P_{i+}(1 - P_{+j}) \right) \\
&= 4 \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_{i+}P_{+j} = 4\tilde{\lambda}^2
\end{aligned} \tag{50}$$

Le MAOR avec pondération arbitraire, défini comme l'écart-type géométrique pondéré de tous les *odds ratios* de la table, est donc égal à :

$$\widetilde{MAOR} = \exp \sqrt{\frac{4}{\sum_{i=1}^I \sum_{j=1}^J P_{i+}(1 - P_{i+})P_{+j}(1 - P_{+j})}} \tilde{\lambda} \tag{51}$$

Comme annoncé plus haut, on voit qu'en remplaçant dans (50) P_{i+} par $1/I$ et P_{+j} par $1/J$, on retrouve la formule du MAOR avec pondération uniforme présentée plus haut :

$$\begin{aligned}
 MAOR &= \exp \sqrt{\frac{4}{\sum_{i=1}^I \sum_{j=1}^J \frac{1}{I} (1 - \frac{1}{I}) \frac{1}{J} (1 - \frac{1}{J})} \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \frac{1}{IJ}} \\
 &= \exp \sqrt{\frac{4}{\sum_{i=1}^I (I-1)/I^2 \sum_{j=1}^J (J-1)/J^2} \lambda^\dagger} \\
 &= \exp \sqrt{\frac{4IJ}{(I-1)(J-1)} \lambda^\dagger}
 \end{aligned} \tag{52}$$

Dans le cas particulier d'une table de dimensions 2×2 , le MAOR avec pondération arbitraire se ramène à l'*odds ratio*. En effet, d'après (6) et (7) (p. 528), on trouve dans le cas d'une table 2×2 (Goodman, 1996, p. 411) :

$$\tilde{\lambda} = \log[(P_{11}P_{22})/(P_{12}P_{21})] \sqrt{P_{1+}P_{2+}P_{+1}P_{+2}} \tag{53}$$

Or, dans une telle table, le MAOR avec pondération arbitraire s'écrit :

$$\begin{aligned}
 \widetilde{MAOR} &= \exp \sqrt{\frac{4}{\sum_{i=1}^2 \sum_{j=1}^2 P_{i+} (1 - P_{i+}) P_{+j} (1 - P_{+j})} \tilde{\lambda}} \\
 &= \exp \sqrt{\frac{4}{P_{1+}P_{2+}P_{+1}P_{+2} + P_{1+}P_{2+}P_{+2}P_{+1} + P_{2+}P_{1+}P_{+1}P_{+2} + P_{2+}P_{1+}P_{+2}P_{+1}} \tilde{\lambda}} \\
 &= \exp \sqrt{\frac{1}{P_{1+}P_{2+}P_{+1}P_{+2}} \tilde{\lambda}} \\
 &= (P_{11}P_{22})/(P_{12}P_{21}) \\
 &= \theta_{11,22}
 \end{aligned} \tag{54}$$

Il est intéressant de souligner que ce résultat est valable même lorsqu'une pondération marginale est utilisée. Cela se comprend aisément en repartant de la définition que nous avons donnée de cet indicateur : en tant que moyenne (pondérée) des *odds ratios* de la table, le MAOR pondéré est nécessairement égal à l'unique *odds ratio* (ou à son inverse) d'une telle table, qui est lui-même indépendant des marges. Ainsi, les pondérations n'interviennent que pour définir le poids relatif dans la moyenne de chacun des *odds ratios* de la table : en présence d'un seul *odds ratio*, les poids n'interviennent pas. Le MAOR pondéré constitue donc une généralisation de l'*odds ratio* dans laquelle les différentes catégories se voient

attribuer une importance variable ; le MAOR avec pondération uniforme n'est qu'un cas particulier de cette généralisation.

3 Relation avec le modèle UNIDIFF

Le MAOR et le coefficient d'association intrinsèque sont directement liés à l'association représentée par le modèle log-multiplicatif à effet de couche (Xie, 1992), plus connu sous le nom d'UNIDIFF (Erikson et Goldthorpe, 1992)¹⁰. En effet, les proportions attendues sous ce modèle suivent l'équation, avec F_{ijk} l'effectif prédit par le modèle dans la cellule à l'intersection de la ligne i , de la colonne j et de la couche k :

$$\log F_{ijk} = \lambda^0 + \lambda_i^I + \lambda_j^J + \lambda_k^K + \lambda_{ik}^{IK} + \lambda_{jk}^{JK} + \phi_k \psi_{ij} \quad (55)$$

Pour une couche donnée k , si le modèle est vérifié, les coefficients d'interaction entre lignes et colonnes s'écrivent $\lambda_{ijk} = \phi_k \psi_{ij}$. Le coefficient d'association intrinsèque non pondéré pour la couche k est donc égal à :

$$\lambda_k^2 = \sum_{i=1}^I \sum_{j=1}^J (\phi_k \psi_{ij})^2 = \phi_k^2 \sum_{i=1}^I \sum_{j=1}^J \psi_{ij}^2 \quad (56)$$

Il suit donc que le rapport entre les intensités des associations relatives aux couches k et k' est égal au rapport entre les coefficients d'association intrinsèque respectifs de ces couches :

$$\frac{\lambda_{k'}}{\lambda_k} = \frac{\phi_{k'}}{\phi_k} \quad (57)$$

De ce fait, si on note par l'indice $k = 0$ la couche de référence pour laquelle le coefficient d'effet de couche est défini comme égal à 1, soit $\phi_0 = 1$, on a :

$$\lambda_k = \phi_k \lambda_0 \quad (58)$$

10 Cette relation a déjà été abordée en partie par Leo Goodman et Michael Hout (2001), mais concernant des modèles qui, plus généraux par certains aspects, sont beaucoup plus restrictifs qu'UNIDIFF par d'autres.

Par conséquent, le rapport entre les logarithmes des MAOR avec pondération uniforme correspondant à deux couches est égal au rapport entre leurs coefficients d'effet de couche ϕ_k :

$$\frac{\log MAOR_{k'}}{\log MAOR_k} = \frac{\sqrt{4/[(I-1)(J-1)]}\lambda_{k'}}{\sqrt{4/[(I-1)(J-1)]}\lambda_k} = \frac{\lambda_{k'}}{\lambda_k} = \frac{\phi_{k'}}{\phi_k} \quad (59)$$

soit :

$$MAOR_{k'} = (MAOR_k)^{\frac{\phi_{k'}}{\phi_k}} \quad (60)$$

ou encore, en prenant $k = 0$ comme couche de référence :

$$MAOR_{k'} = (MAOR_0)^{\phi_{k'}} \quad (61)$$

On peut facilement vérifier par la même démarche que cette propriété est vérifiée avec le MAOR avec pondération arbitraire, du moment que les pondérations sont indépendantes de la couche considérée : c'est le cas notamment de la pondération par les marges de l'ensemble de la table (par opposition aux marges de la seule couche considérée), qui constitue une alternative intéressante à la pondération uniforme dans l'optique d'étudier les variations entre les couches de l'intensité de l'association indépendamment des marges de la table.

Présenter les *odds ratios* absolus moyens ou les coefficients d'association intrinsèque relatifs aux différentes couches permet donc de comparer celles-ci de la même manière qu'en utilisant les coefficients d'effet de couche ϕ_k , et le cas échéant de recalculer ces coefficients. Les deux premières mesures présentent l'avantage supplémentaire de donner la mesure de l'ampleur absolue de l'association, alors que les coefficients d'effet de couche ne permettent qu'une lecture relativement à la couche de référence.

Finalement, notons que dans le cas où le modèle UNIDIFF ne décrit pas correctement les données, le MAOR et le coefficient d'association intrinsèque peuvent être utilisés pour mesurer l'intensité de l'association relative aux différentes couches sans faire l'hypothèse que la structure de l'association est homogène entre les couches. En ce sens, ces indices constituent des généralisations de la mesure fournie par le coefficient d'effet de couche du modèle UNIDIFF. Cette généralisation peut être réalisée soit en calculant la valeur de ces

indices directement à partir des données observées, soit en les combinant avec des modèles plus complexes qu'UNIDIFF, comme le modèle à forme de régression (Goodman et Hout, 1998), ou le modèle d'association lignes-colonnes avec effet de couche RC(M)-L que nous développons dans la section suivante.

4 Relation avec les modèles d'association lignes-colonnes

Le coefficient d'association intrinsèque a été élaboré par Goodman dans le cadre des modèles d'association (Goodman, 1981, 1985, 1986 ; Becker et Clogg, 1989 ; Clogg et Shihadeh, 1994 ; Wong, 2010) : il est donc directement relié à ces modèles, et le MAOR hérite de cette relation étroite. Sous le modèle log-multiplicatif d'association lignes-colonnes RC (aussi parfois appelé modèle de type II de Goodman), les proportions attendues suivent l'équation, avec F_{ij} l'effectif prédit par le modèle dans la cellule à l'intersection de la ligne i et de la colonne j :

$$\log F_{ij} = \lambda^0 + \lambda_i^I + \lambda_j^J + \sum_{m=1}^M \phi_m \mu_{im} v_{jm} \quad (62)$$

Dans cette équation, ϕ_m est le coefficient d'association intrinsèque de la dimension m , tandis que les μ_{im} et v_{jm} sont les scores correspondant aux catégories respectivement en lignes et en colonnes. Par convention, et sans perte de généralité, ϕ_m est toujours choisi comme positif (moyennant le cas échéant un changement de signe des μ_{im} ou des v_{jm}). Ces coefficients sont rendus identifiables à l'aide des contraintes suivantes (contraintes de position, d'échelle et de non corrélation entre les dimensions) :

$$\begin{aligned} \sum_{i=1}^I \mu_{im} &= \sum_{j=1}^J v_{jm} = 0, \\ \sum_{i=1}^I \mu_{im}^2 &= \sum_{j=1}^J v_{jm}^2 = 1, \\ \sum_{i=1}^I \mu_{im} \mu_{im'} &= \sum_{j=1}^J v_{jm} v_{jm'} = 0 \text{ pour tout } m \neq m' \end{aligned} \quad (63)$$

Dans la version pondérée, l'équation du modèle est :

$$\log F_{ij} = \tilde{\lambda}^0 + \tilde{\lambda}_i^I + \tilde{\lambda}_j^J + \sum_{m=1}^M \tilde{\phi}_m \tilde{\mu}_{im} \tilde{\nu}_{jm} \quad (64)$$

$\tilde{\phi}_m$, $\tilde{\mu}_{im}$ et $\tilde{\nu}_{jm}$ sont définis de manière similaire, mais avec les contraintes d'identification pondérées suivantes :

$$\begin{aligned} \sum_{i=1}^I \tilde{\mu}_{im} P_{i+} &= \sum_{j=1}^J \tilde{\nu}_{jm} P_{+j} = 0, \\ \sum_{i=1}^I \tilde{\mu}_{im}^2 P_{i+} &= \sum_{j=1}^J \tilde{\nu}_{jm}^2 P_{+j} = 1, \\ \sum_{i=1}^I \tilde{\mu}_{im} \tilde{\mu}_{im'} P_{i+} &= \sum_{j=1}^J \tilde{\nu}_{jm} \tilde{\nu}_{jm'} P_{+j} = 0 \text{ pour tout } m \neq m' \end{aligned} \quad (65)$$

4.1 Le coefficient d'association intrinsèque

Dans un modèle d'association, l'importance d'une dimension se mesure au coefficient d'association intrinsèque correspondant, généralement noté ϕ_m pour la dimension m . Ce coefficient est l'équivalent direct du coefficient éponyme noté jusqu'ici λ , mais calculé à partir de la composante de l'interaction totale entre lignes et colonnes qui revient à la dimension considérée. En effet, d'après l'équation (62), on voit que sous le modèle RC(M), le coefficient d'interaction entre lignes et colonnes est égal à :

$$\lambda_{ij} = \sum_{m=1}^M \phi_m \mu_{im} \nu_{jm} \quad (66)$$

Chaque dimension contribue ainsi à l'interaction à hauteur de $\phi_m \mu_{im} \nu_{jm}$ (une valeur qui peut être négative aussi bien que positive pour une cellule et une dimension donnée). Or, si l'on reprend l'équation (10) (p. 529) définissant le coefficient d'association intrinsèque d'ensemble λ , mais en remplaçant le terme λ_{ij} par $\phi_m \mu_{im} \nu_{jm}$ de manière à ne tenir compte que de l'apport de la dimension m , on obtient, d'après (62) et (63) :

$$\lambda_m^2 = \sum_{i=1}^I \sum_{j=1}^J (\phi_m \mu_{im} \nu_{jm})^2 = \phi_m^2 \sum_{i=1}^I \mu_{im}^2 \sum_{j=1}^J \nu_{jm}^2 = \phi_m^2 \quad (67)$$

Et dans la version pondérée, en reprenant (11) (p. 529) d'après (64) et (65) :

$$\tilde{\lambda}_m^2 = \sum_{i=1}^I \sum_{j=1}^J (\tilde{\phi}_m \tilde{\mu}_{im} \tilde{\nu}_{jm})^2 P_{i+P+j} = \tilde{\phi}_m^2 \sum_{i=1}^I \tilde{\mu}_{im}^2 P_{i+} \sum_{j=1}^J \tilde{\nu}_{jm}^2 P_{+j} = \tilde{\phi}_m^2 \quad (68)$$

On voit que le coefficient d'association intrinsèque ϕ_m (respectivement $\tilde{\phi}_m$) est bien l'équivalent, appliqué à une dimension particulière, du coefficient éponyme λ (respectivement $\tilde{\lambda}$) défini dans la première section concernant l'association totale. Cette relation va au-delà d'une simple analogie : le coefficient d'association intrinsèque d'ensemble est égal à la norme euclidienne des coefficients correspondant à chaque dimension. En effet, d'après (10) (p. 529) et (63) :

$$\begin{aligned} \lambda^2 &= \sum_{i=1}^I \sum_{j=1}^J \left(\sum_{m=1}^M \phi_m \mu_{im} \nu_{jm} \right)^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M (\phi_m \mu_{im} \nu_{jm})^2 + \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M \sum_{m' \neq m} \phi_m \phi_{m'} \mu_{im} \mu_{im'} \nu_{jm} \nu_{jm'} \\ &= \sum_{m=1}^M \lambda_m^2 + \sum_{m=1}^M \sum_{m' \neq m} \phi_m \phi_{m'} \sum_{i=1}^I \mu_{im} \mu_{im'} \sum_{j=1}^J \nu_{jm} \nu_{jm'} \\ &= \sum_{m=1}^M \lambda_m^2 \end{aligned} \quad (69)$$

De même, dans la version pondérée, d'après (11) (p. 529) et (65) :

$$\begin{aligned} \tilde{\lambda}^2 &= \sum_{i=1}^I \sum_{j=1}^J \left(\sum_{m=1}^M \tilde{\phi}_m \tilde{\mu}_{im} \tilde{\nu}_{jm} \right)^2 P_{i+P+j} \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M (\tilde{\phi}_m \tilde{\mu}_{im} \tilde{\nu}_{jm})^2 P_{i+P+j} + \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M \sum_{m' \neq m} \tilde{\phi}_m \tilde{\phi}_{m'} \tilde{\mu}_{im} \tilde{\mu}_{im'} \tilde{\nu}_{jm} \tilde{\nu}_{jm'} P_{i+P+j} \\ &= \sum_{m=1}^M \tilde{\lambda}_m^2 + \sum_{m=1}^M \sum_{m' \neq m} \tilde{\phi}_m \tilde{\phi}_{m'} \sum_{i=1}^I \tilde{\mu}_{im} \tilde{\mu}_{im'} P_{i+} \sum_{j=1}^J \tilde{\nu}_{jm} \tilde{\nu}_{jm'} P_{+j} \\ &= \sum_{m=1}^M \tilde{\lambda}_m^2 \end{aligned} \quad (70)$$

Ainsi, le coefficient d'association intrinsèque d'ensemble (pondéré ou non) est égal à la norme euclidienne des coefficients d'association intrinsèque correspondant à chaque dimension du modèle. Un modèle d'association revient donc à décomposer l'association de la table en une série de dimensions d'importance décroissante. Cette décomposition est

valable aussi bien lorsque le modèle est saturé que lorsqu'il ne l'est pas, mais qu'il s'ajuste correctement aux données. Les modèles d'association sont donc à la tradition de l'*odds ratio* ce que sont l'analyse de corrélation ou l'analyse des correspondances à la tradition pearsonienne (Goodman, 1985, 1986, 1991, 1996 ; Gilula et Haberman, 1986).

4.2 Le MAOR dans les modèles RC(M)

Cette décomposition peut aussi être interprétée en termes d'*odds ratio* absolu moyen. En reprenant la définition (40) (p. 538), on peut définir le MAOR avec pondération uniforme correspondant à une dimension donnée à partir du coefficient d'association intrinsèque de cette dimension :

$$MAOR_m = \exp \sqrt{\frac{4}{(I-1)(J-1)}} \lambda_m \quad (71)$$

De même que pour la version pondérée, définie en (51) (p. 543) :

$$\widetilde{MAOR}_m = \exp \sqrt{\frac{4}{\sum_{i=1}^I \sum_{j=1}^J P_{i+}(1-P_{i+})P_{+j}(1-P_{+j})}} \tilde{\lambda}_m \quad (72)$$

Ce dernier s'interprète comme l'*odds ratio* absolu moyen qui serait observé si seule cette dimension était prise en compte, et fournit donc une mesure de l'intensité des contrastes entre catégories sur la dimension considérée. Néanmoins, la relation entre les MAOR ainsi définis pour chaque dimension du modèle et le MAOR d'ensemble ne s'écrit pas sous une forme simple¹¹. La connaissance des MAOR relatifs à chaque dimension suffit en revanche à calculer le MAOR d'ensemble, ce qui en fait des indicateurs aussi riches que le coefficient d'association intrinsèque.

4.3 Extension au modèle RC(M)-L

Les modèles RC(M)-L (Clogg, 1982 ; Wong, 2010) sont une extension des modèles RC(M) à des tables de trois dimensions : le coefficient d'association intrinsèque et/ou les scores sont autorisés à varier selon les couches (troisième dimension de la table). Une version de ce modèle postule que l'association est identique pour toutes les couches (scores et coefficients d'association intrinsèque homogènes) ; une seconde, qu'elle est complètement différente

¹¹ Une telle propriété serait obtenue en utilisant une norme 1 (valeur absolue) plutôt qu'une norme 2 (carré).

entre les couches (scores et coefficients d'association intrinsèque hétérogènes) : ces deux versions peuvent être ramenées, la première à un seul modèle RC(M), la seconde à autant de modèles RC(M) que de couches.

Seule la troisième version du modèle RC(M)-L exige une extension de l'approche développée jusqu'ici. Cette version suppose que les scores sont homogènes mais que les coefficients d'association intrinsèque sont hétérogènes entre les couches. Son équation s'écrit, avec F_{ijk} l'effectif prédit par le modèle dans la cellule à l'intersection de la ligne i , de la colonne j et de la couche k :

$$\log F_{ijk} = \lambda^0 + \lambda_i^I + \lambda_j^J + \lambda_k^K + \lambda_{ik}^{IK} + \lambda_{jk}^{JK} + \sum_{m=1}^M \phi_{mk} \mu_{im} \nu_{jm} \quad (73)$$

ou, dans sa version pondérée :

$$\log F_{ijk} = \lambda^0 + \tilde{\lambda}_i^I + \tilde{\lambda}_j^J + \tilde{\lambda}_k^K + \tilde{\lambda}_{ik}^{IK} + \tilde{\lambda}_{jk}^{JK} + \sum_{m=1}^M \tilde{\phi}_{mk} \tilde{\mu}_{im} \tilde{\nu}_{jm} \quad (74)$$

Seules les deux premières contraintes des équations (63) et (65) (p. 547), s'appliquant aux scores, sont requises : les contraintes interdimensionnelles ne peuvent plus être appliquées, et il existe généralement une corrélation non nulle entre les scores des différentes dimensions. De ce fait, la logique des équations (69) et (70) (p. 549) ne peut plus être suivie. La relation entre les coefficients d'association intrinsèque relatifs à chaque dimension et le coefficient d'association intrinsèque d'ensemble n'est plus une simple somme : elle doit tenir compte de la corrélation entre les dimensions. Dans les équations (69) et (70) (p. 549) respectivement, les termes :

$$\sum_{m=1}^M \sum_{m' \neq m} \phi_m \phi_{m'} \sum_{i=1}^I \mu_{im} \mu_{im'} \sum_{j=1}^J \nu_{jm} \nu_{jm'} \quad (75)$$

et

$$\sum_{m=1}^M \sum_{m' \neq m} \tilde{\phi}_m \tilde{\phi}_{m'} \sum_{i=1}^I \tilde{\mu}_{im} \tilde{\mu}_{im'} P_{i+} \sum_{j=1}^J \tilde{\nu}_{jm} \tilde{\nu}_{jm'} P_{+j} \quad (76)$$

qui correspondent à la somme des produits des coefficients d'association intrinsèque et des corrélations entre les scores respectivement en lignes et en colonnes des dimensions prises deux à deux, ne sont pas en général égaux à 0. Ainsi, l'intensité de l'association sur

une couche donnée dépend de la corrélation positive, négative ou nulle, entre les scores, ainsi que des intensités respectives des différentes dimensions. Intuitivement, une dimension peut contrarier l'association représentée par une autre si elle est assez forte et si les deux dimensions ont des scores suffisamment différents. Soulignons en outre que les coefficients d'association intrinsèque peuvent ici être négatifs, ce qui revient en fait à inverser le signe des scores en lignes ou en colonnes, et donc à inverser le sens du lien.

Malgré cette complexité accrue, qui provient de la richesse du modèle RC(M)-L, le coefficient d'association intrinsèque aussi bien que le MAOR peuvent toujours être calculés séparément, d'un côté pour l'association d'ensemble, de l'autre pour chacune des dimensions. Seule la relation entre les deux quantités cesse de suivre une formule d'interprétation simple ; mesurer l'intensité d'une dimension prise séparément demeure cependant d'un grand intérêt pour l'analyse.

Pour finir, notons que, comme pour le modèle UNIDIFF, l'analyse des différences dans l'association indépendamment des variations des marges entre les couches incite à retenir soit une pondération uniforme, soit une pondération par les marges en lignes et en colonnes de l'ensemble de la table (plutôt que par les marges de chaque couche).

5 Composantes symétrique et anti-symétrique de l'association

Le coefficient d'association intrinsèque et le MAOR peuvent toujours être décomposés en deux parties, symétrique et anti-symétrique. En effet, dans une table carrée de dimensions $I \times I$, la décomposition suivante des coefficients d'interaction peut toujours être réalisée (van der Heijden et Mooijaart, 1995), que ceux-ci correspondent aux données observées (modèle saturé) ou à un modèle plus parcimonieux :

$$\lambda_{ij} = s_{ij} + a_{ij} \quad \text{avec } s_{ij} = \frac{\lambda_{ij} + \lambda_{ji}}{2} \quad \text{et} \quad \sum_{i=1}^I s_{ij} = \sum_{j=1}^J s_{ij} = 0$$

$$\text{avec } a_{ij} = \frac{\lambda_{ij} - \lambda_{ji}}{2} \quad \text{et} \quad \sum_{i=1}^I a_{ij} = \sum_{j=1}^J a_{ij} = 0 \quad (77)$$

Dans cette équation, s_{ij} est la composante symétrique de l'association, et a_{ij} la composante anti-symétrique. En effet, on voit que, suivant la définition de ces deux quantités, $s_{ij} = s_{ji}$ et $a_{ij} = -a_{ji}$.

En recourant à des poids arbitraires symétriques entre lignes et colonnes notés P_i , on obtient l'équivalent pondéré de l'équation ci-dessus :

$$\begin{aligned} \tilde{\lambda}_{ij} = \tilde{s}_{ij} + \tilde{a}_{ij} \quad \text{avec } \tilde{s}_{ij} = \frac{\tilde{\lambda}_{ij} + \tilde{\lambda}_{ji}}{2} \quad \text{et} \quad \sum_{i=1}^I \tilde{s}_{ij} P_i P_j = \sum_{j=1}^J \tilde{s}_{ij} P_i P_j = 0 \\ \text{avec } \tilde{a}_{ij} = \frac{\tilde{\lambda}_{ij} - \tilde{\lambda}_{ji}}{2} \quad \text{et} \quad \sum_{i=1}^I \tilde{a}_{ij} P_i P_j = \sum_{j=1}^J \tilde{a}_{ij} P_i P_j = 0 \end{aligned} \quad (78)$$

En particulier, une pondération par la moyenne des marges en lignes et en colonnes constitue ici l'équivalent de la pondération marginale utilisée plus haut :

$$P_i = \frac{P_{i+} + P_{+i}}{2} \quad (79)$$

5.1 Le coefficient d'association intrinsèque

D'après ces formules, le coefficient d'association intrinsèque se décompose de même en deux variantes représentant chacun une source de l'association. La première, notée λ^S , mesure l'intensité de l'association symétrique ; la seconde, notée λ^A , celle de l'association anti-symétrique :

$$\begin{aligned} \lambda^2 &= \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J (s_{ij} + a_{ij})^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J s_{ij}^2 + \sum_{i=1}^I \sum_{j=1}^J a_{ij}^2 + 2 \sum_{i=1}^I \sum_{i < j} s_{ij} a_{ij} + 2 \sum_{i=1}^I \sum_{j < i} s_{ij} a_{ij} \\ &= \sum_{i=1}^I \sum_{j=1}^J s_{ij}^2 + \sum_{i=1}^I \sum_{j=1}^J a_{ij}^2 + 2 \sum_{i=1}^I \sum_{i < j} s_{ij} a_{ij} - 2 \sum_{i=1}^I \sum_{i < j} s_{ij} a_{ij} \\ &= \sum_{i=1}^I \sum_{j=1}^J s_{ij}^2 + \sum_{i=1}^I \sum_{j=1}^J a_{ij}^2 \\ &= (\lambda^S)^2 + (\lambda^A)^2 \end{aligned} \quad (80)$$

De même, la version pondérée de cette équation s'écrit :

$$\begin{aligned}
\tilde{\lambda}^2 &= \sum_{i=1}^I \sum_{j=1}^J \tilde{\lambda}_{ij}^2 P_i P_j \\
&= \sum_{i=1}^I \sum_{j=1}^J (\tilde{s}_{ij} + \tilde{a}_{ij})^2 P_i P_j \\
&= \sum_{i=1}^I \sum_{j=1}^J \tilde{s}_{ij}^2 P_i P_j + \sum_{i=1}^I \sum_{j=1}^J \tilde{a}_{ij}^2 P_i P_j + 2 \sum_{i=1}^I \sum_{i < j} \tilde{s}_{ij} \tilde{a}_{ij} P_i P_j + 2 \sum_{i=1}^I \sum_{j < i} \tilde{s}_{ij} \tilde{a}_{ij} P_i P_j \\
&= \sum_{i=1}^I \sum_{j=1}^J \tilde{s}_{ij}^2 P_i P_j + \sum_{i=1}^I \sum_{j=1}^J \tilde{a}_{ij}^2 P_i P_j + 2 \sum_{i=1}^I \sum_{i < j} \tilde{s}_{ij} \tilde{a}_{ij} P_i P_j - 2 \sum_{i=1}^I \sum_{i < j} \tilde{s}_{ij} \tilde{a}_{ij} P_i P_j \\
&= \sum_{i=1}^I \sum_{j=1}^J \tilde{s}_{ij}^2 P_i P_j + \sum_{i=1}^I \sum_{j=1}^J \tilde{a}_{ij}^2 P_i P_j \\
&= (\tilde{\lambda}^S)^2 + (\tilde{\lambda}^A)^2
\end{aligned} \tag{81}$$

On obtient en particulier à partir de cette dernière formule la décomposition suivant une pondération uniforme (qui est par définition symétrique dans une table carrée) :

$$(\lambda^\dagger)^2 = (\lambda^{\dagger S})^2 + (\lambda^{\dagger A})^2 \tag{82}$$

5.2 Le MAOR symétrique et anti-symétrique

Des versions symétriques et anti-symétriques du MAOR peuvent être définies en repartant des *odds ratios*. En effet, on peut définir d'une part un *odds ratio* symétrique Θ^S égal à la moyenne géométrique d'un *odds ratio* et de son symétrique, d'après les équations (23) (p. 532) et (77) (p. 552) :

$$\begin{aligned}
 \Theta_{ij,i'j'}^S &= \sqrt{\theta_{ij,i'j'}\theta_{j'i',ji'}} \\
 &= \sqrt{\exp[\lambda_{ij} + \lambda_{i'j'} - \lambda_{i'j} - \lambda_{ij'}] \times \exp[\lambda_{ji} + \lambda_{j'i'} - \lambda_{j'i} - \lambda_{j'i}]} \\
 &= \exp\left[\frac{1}{2}(\lambda_{ij} + \lambda_{i'j'} - \lambda_{i'j} - \lambda_{ij'}) + \frac{1}{2}(\lambda_{ji} + \lambda_{j'i'} - \lambda_{j'i} - \lambda_{j'i})\right] \\
 &= \exp\left[\frac{1}{2}((s_{ij} + a_{ij}) + (s_{i'j'} + a_{i'j'}) - (s_{i'j} + a_{i'j}) - (s_{ij'} + a_{ij'}))\right. \\
 &\quad \left. + \frac{1}{2}((s_{ji} + a_{ji}) + (s_{j'i'} + a_{j'i'}) - (s_{j'i} + a_{j'i}) - (s_{j'i'} + a_{j'i'}))\right] \\
 &= \exp\left[\frac{1}{2}((s_{ij} + a_{ij}) + (s_{i'j'} + a_{i'j'}) - (s_{i'j} + a_{i'j}) - (s_{ij'} + a_{ij'}))\right. \\
 &\quad \left. + \frac{1}{2}((s_{ij} - a_{ij}) + (s_{i'j'} - a_{i'j'}) - (s_{i'j} - a_{i'j}) - (s_{ij'} - a_{ij'}))\right] \\
 &= \exp[(s_{ij} + s_{i'j'} - s_{i'j} - s_{ij'})]
 \end{aligned} \tag{83}$$

On voit que cet *odds ratio* s'exprime uniquement à partir des coefficients d'association symétrique, suivant une expression de même forme que l'*odds ratio* classique.

D'autre part, un *odds ratio* anti-symétrique Θ^A peut être défini comme la moyenne géométrique d'un *odds ratio* et de l'inverse de son symétrique. De manière assez intuitive, il exprime le rapport entre les intensités de l'association observées de part et d'autre de la diagonale de la table (Yamaguchi, 1990).

$$\begin{aligned}
\Theta_{ij,i'j'}^A &= \sqrt{\frac{\theta_{ij,i'j'}}{\theta_{j'i,j'i'}}} \\
&= \sqrt{\frac{\exp[\lambda_{ij} + \lambda_{i'j'} - \lambda_{i'j} - \lambda_{ij'}]}{\exp[\lambda_{ji} + \lambda_{j'i'} - \lambda_{j'i} - \lambda_{j'i}]}} \\
&= \exp\left[\frac{1}{2}(\lambda_{ij} + \lambda_{i'j'} - \lambda_{i'j} - \lambda_{ij'}) - \frac{1}{2}(\lambda_{ji} + \lambda_{j'i'} - \lambda_{j'i} - \lambda_{j'i})\right] \\
&= \exp\left[\frac{1}{2}((s_{ij} + a_{ij}) + (s_{i'j'} + a_{i'j'}) - (s_{i'j} + a_{i'j}) - (s_{ij'} + a_{ij'}))\right. \\
&\quad \left. - \frac{1}{2}((s_{ji} + a_{ji}) + (s_{j'i'} + a_{j'i'}) - (s_{j'i} + a_{j'i}) - (s_{j'i'} + a_{j'i'}))\right] \\
&= \exp\left[\frac{1}{2}((s_{ij} + a_{ij}) + (s_{i'j'} + a_{i'j'}) - (s_{i'j} + a_{i'j}) - (s_{ij'} + a_{ij'}))\right. \\
&\quad \left. - \frac{1}{2}((s_{ij} - a_{ij}) + (s_{i'j'} - a_{i'j'}) - (s_{i'j} - a_{i'j}) - (s_{ij'} - a_{ij'}))\right] \\
&= \exp[(a_{ij} + a_{i'j'} - a_{i'j} - a_{ij'})]
\end{aligned} \tag{84}$$

Cet *odds ratio* s'exprime uniquement à partir des coefficients d'association anti-symétrique, sous la même forme que le précédent.

Ainsi, en appliquant la démarche par laquelle nous avons dérivé la formule du MAOR avec pondération uniforme à partir de l'équation 32 (p. 536), mais en partant des *odds ratios* symétriques ou anti-symétriques, on peut élaborer deux variantes éponymes du MAOR¹². Tout d'abord, une variante symétrique :

$$MAOR^S = \exp \sqrt{\frac{4}{(I-1)(J-1)}} \lambda^S = \exp \sqrt{\frac{4IJ}{(I-1)(J-1)}} \lambda^{\dagger S} \tag{85}$$

et une version anti-symétrique :

$$MAOR^A = \exp \sqrt{\frac{4}{(I-1)(J-1)}} \lambda^A = \exp \sqrt{\frac{4IJ}{(I-1)(J-1)}} \lambda^{\dagger A} \tag{86}$$

Néanmoins, ici encore, la relation entre ces deux variantes du MAOR et le MAOR d'ensemble ne s'exprime pas de manière simple, bien que les deux premières soient suffisantes pour recalculer ce dernier.

12 En effet, les *odds ratios* symétrique et anti-symétrique que nous venons de définir partagent les mêmes contraintes en lignes et en colonnes que les *odds ratios*.

Finalement, on obtient les équivalents avec pondération arbitraire de ces indicateurs. En effet, suivant la même démarche qu'aux équations (83) et (84), mais à partir des équations (29) (p. 534) et (78) (p. 553), on obtient pour l'*odds ratio* symétrique :

$$\Theta_{ij,i'j'}^S = \exp \left[\left(\tilde{s}_{ij} + \tilde{s}_{i'j'} - \tilde{s}_{i'j} - \tilde{s}_{ij'} \right) \right] \quad (87)$$

et pour l'*odds ratio* anti-symétrique :

$$\Theta_{ij,i'j'}^A = \exp \left[\left(\tilde{a}_{ij} + \tilde{a}_{i'j'} - \tilde{a}_{i'j} - \tilde{a}_{ij'} \right) \right] \quad (88)$$

En suivant la démarche développée à partir de l'équation (42) (p. 540), mais avec des poids symétriques, on aboutit alors aux égalités suivantes, pour le MAOR symétrique :

$$\widetilde{MAOR}^S = \exp \sqrt{\frac{4}{\sum_{i=1}^I \sum_{j=1}^J P_i(1-P_i)P_j(1-P_j)}} \tilde{\lambda}^S \quad (89)$$

et pour le MAOR anti-symétrique :

$$\widetilde{MAOR}^A = \exp \sqrt{\frac{4}{\sum_{i=1}^I \sum_{j=1}^J P_i(1-P_i)P_j(1-P_j)}} \tilde{\lambda}^A \quad (90)$$

Ces formules permettent de retrouver le MAOR avec pondération uniforme déjà présenté en (85) et (86).

5.3 Le modèle d'association lignes-colonnes symétrique

Ces formules générales s'appliquent aussi bien à l'association observée dans des données brutes qu'à celle décrite par des modèles. En effet, la composante symétrique de l'association peut être modélisée par diverses spécifications. La plus détaillée d'entre elles, qui décrit parfaitement l'association symétrique, est incluse dans le modèle de quasi-symétrie (Causinus, 1965) : elle consiste à estimer un paramètre pour chaque paire $s_{ij} = s_{ji}$. Mais des spécifications plus parcimonieuses sont possibles, dont notamment une variante du modèle RC(M) dans laquelle les scores des lignes et des colonnes sont contraints à être égaux. Ce modèle suit l'équation, avec F_{ij} l'effectif prédit par le modèle dans la cellule à l'intersection de la ligne i et de la colonne j :

$$\log F_{ij} = \lambda^0 + \lambda_i^I + \lambda_j^J + \sum_{m=1}^M \phi_m^S \mu_{im} \mu_{jm} \quad (91)$$

et dans sa version pondérée :

$$\log F_{ij} = \tilde{\lambda}^0 + \tilde{\lambda}_i^I + \tilde{\lambda}_j^J + \sum_{m=1}^M \tilde{\phi}_m^S \tilde{\mu}_{im} \tilde{\mu}_{jm} \quad (92)$$

Les contraintes décrites plus haut aux équations (63) et (65) (p. 547) s'appliquent aux μ .

On voit à ces équations qu'un tel modèle décrit uniquement l'association symétrique, les scores des lignes et des colonnes étant interchangeable. Par conséquent, la composante a_{ij} est nulle, et le coefficient d'association intrinsèque est égal à sa seule composante symétrique. Les démonstrations établies à propos du modèle RC(M) asymétrique peuvent être directement adaptées à ce cas particulier.

Cette spécification, comme le modèle de quasi-symétrie, peut être combinée avec une spécification de l'association anti-symétrique.

5.4 Le modèle d'association lignes-colonnes anti-symétrique

De même que la composante symétrique de l'association, la composante anti-symétrique peut être modélisée de plusieurs manières. La plus détaillée, qui reconstitue parfaitement les données, consiste à estimer un paramètre pour chaque paire $a_{ij} = -a_{ji}$, à l'exception de la diagonale de la table pour laquelle $a_{ii} = 0$ par définition. Mais il est aussi possible d'étendre la démarche des modèles d'association en suivant le principe de la décomposition de Gower : c'est ce que réalise le modèle d'association lignes-colonnes anti-symétrique de Peter van der Heijden et Ab Mooijaart (1995). Sous ce modèle, les proportions attendues suivent l'équation, avec F_{ij} l'effectif prédit par le modèle dans la cellule à l'intersection de la ligne i et de la colonne j :

$$\log F_{ij} = \lambda^0 + \lambda_i^I + \lambda_j^J + s_{ij} + \sum_{m=1}^M \phi_m^A (v_{im} \mu_{jm} - \mu_{im} v_{jm}) \quad (93)$$

Dans cette équation, ϕ_m^A est le coefficient d'association intrinsèque anti-symétrique de la paire de dimensions m , tandis que les μ_{im} et v_{jm} sont les scores correspondant aux catégories respectivement sur la première et la seconde dimension de la paire de dimensions m : on parle dans ce modèle de paire de dimensions (ou encore de « bimension »), puisque chaque valeur de m donne lieu à deux séries de scores. s_{ij} est une spécification symétrique de

l'association, telle que celles décrites dans la section précédente. Les contraintes suivantes, équivalentes à celles appliquées au modèle RC(M), s'appliquent :

$$\begin{aligned}
\sum_{i=1}^I \mu_{im} &= \sum_{i=1}^I v_{im} = 0, \\
\sum_{i=1}^I \mu_{im}^2 &= \sum_{i=1}^I v_{im}^2 = 1, \\
\sum_{i=1}^I \mu_{im} v_{im} &= 0 \text{ pour tout } m, \\
\sum_{i=1}^I \mu_{im} \mu_{im'} &= \sum_{i=1}^I v_{im} v_{im'} = \sum_{i=1}^I \mu_{im} v_{im'} = 0 \text{ pour tout } m \neq m'
\end{aligned} \tag{94}$$

Dans sa version pondérée, l'équation du modèle est :

$$\log F_{ij} = \tilde{\lambda}^0 + \tilde{\lambda}_i^I + \tilde{\lambda}_j^J + \tilde{s}_{ij} + \sum_{m=1}^M \tilde{\phi}_m^A (\tilde{v}_{im} \tilde{\mu}_{jm} - \tilde{\mu}_{im} \tilde{v}_{jm}) \tag{95}$$

$\tilde{\phi}_m^A$, $\tilde{\mu}_{im}$ et \tilde{v}_{jm} sont définis de manière similaire à la version non pondérée, mais avec les contraintes d'identification pondérées suivantes :

$$\begin{aligned}
\sum_{i=1}^I \tilde{\mu}_{im} P_i &= \sum_{i=1}^I \tilde{v}_{im} P_i = 0, \\
\sum_{i=1}^I \tilde{\mu}_{im}^2 P_i &= \sum_{i=1}^I \tilde{v}_{im}^2 P_i = 1, \\
\sum_{i=1}^I \tilde{\mu}_{im} \tilde{v}_{im} P_i &= 0 \text{ pour tout } m, \\
\sum_{i=1}^I \tilde{\mu}_{im} \tilde{\mu}_{im'} P_i &= \sum_{i=1}^I \tilde{v}_{im} \tilde{v}_{im'} P_i = \sum_{i=1}^I \tilde{\mu}_{im} \tilde{v}_{im'} P_i = 0 \text{ pour tout } m \neq m'
\end{aligned} \tag{96}$$

Comme le modèle RC(M), le modèle d'association anti-symétrique est très directement lié au coefficient d'association intrinsèque, et par conséquent au MAOR, dans leurs composantes anti-symétriques. En effet, dans ce modèle, l'importance d'une paire de dimensions se mesure au coefficient d'association intrinsèque anti-symétrique

correspondant, que nous noterons ϕ_m^A pour la paire de dimensions m . Ce coefficient est l'équivalent direct du coefficient λ^A et de ses équivalents pondérés définis en (80), (81) et (82) (p. 554), mais calculé à partir de la composante de l'interaction totale entre lignes et colonnes qui revient à la paire de dimensions considérées. En effet, d'après l'équation (93), on voit que sous le modèle d'association anti-symétrique, le coefficient d'interaction anti-symétrique a_{ij} est égal à :

$$a_{ij} = \sum_{m=1}^M \phi_m^A (v_{im}\mu_{jm} - \mu_{im}v_{jm}) \quad (97)$$

Or, si l'on reprend l'équation (80) (p. 553) définissant le coefficient d'association intrinsèque anti-symétrique λ^A , mais en remplaçant le terme a_{ij} par $\phi_m^A (v_{im}\mu_{jm} - \mu_{im}v_{jm})$ de manière à ne tenir compte que de l'apport de la paire de dimensions m , on obtient :

$$\begin{aligned} (\lambda_m^A)^2 &= \sum_{i=1}^I \sum_{j=1}^J [\phi_m^A (v_{im}\mu_{jm} - \mu_{im}v_{jm})]^2 \\ &= (\phi_m^A)^2 \sum_{i=1}^I \sum_{j=1}^J (v_{im}\mu_{jm} - \mu_{im}v_{jm})^2 \\ &= (\phi_m^A)^2 \left[\sum_{i=1}^I \sum_{j=1}^J v_{im}^2 \mu_{jm}^2 + \sum_{i=1}^I \sum_{j=1}^J \mu_{im}^2 v_{jm}^2 - 2 \sum_{i=1}^I \sum_{j=1}^J \mu_{im} \mu_{jm} v_{im} v_{jm} \right] \\ &= 2(\phi_m^A)^2 \sum_{i=1}^I v_{im}^2 \sum_{j=1}^J \mu_{jm}^2 \\ &= 2(\phi_m^A)^2 \end{aligned} \quad (98)$$

Le même résultat est obtenu concernant sa version pondérée, en reprenant (81) (p. 554) d'après (93) et (94) :

$$\begin{aligned}
(\tilde{\lambda}_m^A)^2 &= \sum_{i=1}^I \sum_{j=1}^J \left[\tilde{\phi}_m^A (\tilde{v}_{im} \tilde{\mu}_{jm} - \tilde{\mu}_{im} \tilde{v}_{jm}) \right]^2 P_i P_j \\
&= (\tilde{\phi}_m^A)^2 \sum_{i=1}^I \sum_{j=1}^J \tilde{v}_{im}^2 \tilde{\mu}_{jm}^2 P_i P_j \\
&\quad + (\tilde{\phi}_m^A)^2 \sum_{i=1}^I \sum_{j=1}^J \tilde{\mu}_{im}^2 \tilde{v}_{jm}^2 P_i P_j \\
&\quad - 2(\tilde{\phi}_m^A)^2 \sum_{i=1}^I \tilde{\mu}_{im} \tilde{v}_{im} P_i \sum_{j=1}^J \tilde{\mu}_{jm} \tilde{v}_{jm} P_j \\
&= 2(\tilde{\phi}_m^A)^2 \sum_{i=1}^I \tilde{v}_{im}^2 P_i \sum_{j=1}^J \tilde{\mu}_{jm}^2 P_j \\
&= 2(\tilde{\phi}_m^A)^2
\end{aligned} \tag{99}$$

Le coefficient d'association intrinsèque anti-symétrique est donc égal au coefficient d'association de la paire de dimensions considérée multiplié par la constante $\sqrt{2}$ qui reflète le fait que deux dimensions sont en jeu, de manière similaire à un modèle RC(2).

On peut en dériver, comme pour le modèle RC(M), un MAOR anti-symétrique correspondant à une paire de dimensions donnée, d'après la définition (40) (p. 538) :

$$MAOR_m^A = \exp \sqrt{\frac{4}{(I-1)(J-1)}} \lambda_m^A \tag{100}$$

De même qu'une version pondérée, définie en (51) (p. 543) :

$$\widetilde{MAOR}_m^A = \exp \sqrt{\frac{4}{\sum_{i=1}^I \sum_{j=1}^J P_i (1-P_i) P_j (1-P_j)}} \tilde{\lambda}_m^A \tag{101}$$

Ce MAOR s'interprète comme l'intensité de l'anti-symétrie (rapport entre un *odds ratio* et son symétrique) qui serait observée si seule la paire de dimensions considérée était prise en compte.

Nous ne développons pas ici la démonstration concernant la relation entre le coefficient d'association intrinsèque anti-symétrique de chaque dimension et celui d'ensemble lorsque

plus d'une paire de dimensions est introduite dans le modèle, étant donné que cette extension n'a à notre connaissance jamais été utilisée.

BIBLIOGRAPHIE

- ABDI H., 2010, « Partial least squares regression and projection on latent structure regression (PLS Regression) », *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 1, p. 97-106.
- AGRESTI A., [1990] 2002, *Categorical Data Analysis*, 2^e édition, New York, Wiley (Probabilities and Statistics).
- AHO K., DERRYBERRY D., PETERSON T., 2014, « Model selection for ecologists: the worldviews of AIC and BIC », *Ecology*, 95, 3, p. 631-636.
- AKAIKE H., 1974, « A new look at the statistical model identification », *IEEE Transactions on Automatic Control*, 19, 6, p. 716-723.
- ALBERONI F., [1979] 1993, *Le choc amoureux*, Paris, Pocket.
- ALBOUY V., BREUIL-GENIER P., 2012, « Démographie et famille: les différences sociales se réduisent-elles ? », *France, portrait social 2012*, Paris, Insee (Insee Références), p. 11-18.
- ALBOUY V., DJIDER Z., MAINGUENÉ A., 2012, « Activité, emploi, salaires et retraites: la convergence des situations entre hommes et femmes s'opère, mais parfois bien lentement », *Regards sur la parité 2012*, Paris, Insee (Insee Références), p. 67-80.
- ALBOUY V., WANECQ T., 2003, « Les inégalités sociales d'accès aux grandes écoles », *Économie et statistique*, 361, p. 27-52.
- AMAR M., 2010, « Les très hauts salaires du secteur privé », *Insee Première*, 1288.
- AMOSSÉ T., 2004, « Professions au féminin: Représentation statistique, construction sociale », *Travail, genre et sociétés*, 11, p. 31-46.
- AMOSSÉ T., CHARDON O., 2006, « Les travailleurs non qualifiés: une nouvelle classe sociale ? », *Économie et statistique*, 393, p. 203-229.
- AMOSSÉ T., DE PERETTI G., 2011, « Hommes et femmes en ménage statistique: une valse à trois temps », *Travail, genre et sociétés*, 26, p. 23.
- ATTIAS-DONFUT C., BARNAY T., 2002, « Investissement en capital humain en fonction des sexes », *Revue économique*, 53, 4, p. 839-861.
- AUDIRAC P.-A., 1982, « Cohabitation et mariage: qui vit avec qui ? », *Économie et statistique*, 145, p. 41-59.
- AXINN W.G., THORNTON A., 1992, « The influence of parental resources on the timing of the transition to marriage », *Social Science Research*, 21, 3, p. 261-285.
- BARBIER J.-C., 2005, « La précarité, une catégorie française à l'épreuve de la comparaison internationale », *Revue française de sociologie*, 46, 2, p. 351-371.
- BARONE C., SCHIZZEROTTO A., 2008, « The application of the ISCED-97 to Italy », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 149-161.

- BATTAGLIOLA F., 1988, *La fin du mariage? Jeunes couples des années 1980*, Paris, Syros/Alternatives.
- BAUDELLOT C., 1989, « L'âge rend-il plus savant? Un exemple de biais de réponse dans les enquêtes », *Les Ménages: mélanges en l'honneur de Jacques Desabie*, Paris, Insee.
- BAUDELLOT C., ESTABLET R., 1992, *Allez les filles!*, Paris, Seuil.
- BAUER D., 2007, « Entre maison, enfant(s) et travail: les diverses formes d'arrangement dans les couples », *Études et résultats*, 570, Paris, DREES.
- BEAUD S., 2002, *80 % au bac... et après? Les enfants de la démocratisation scolaire*, Paris, La Découverte.
- BEAUD S., PIALOUX M., 1999, *Retour sur la condition ouvrière*, Paris, Fayard.
- BEAUJOUAN É., RÉGNIER-LOILIER A., VILLENEUVE-GOKALP C., 2009, « Ni seuls ni en couple. Les relations amoureuses non cohabitantes », dans RÉGNIER-LOILIER A. (dir.), *Portraits de familles. L'enquête Étude des relations familiales et intergénérationnelles*, Paris, Ined (Grandes enquêtes), p. 87-112.
- BECKER G., 1981, *A Treatise on the Family*, Cambridge, MA, Harvard University Press.
- BECKER M.P., CLOGG C.C., 1989, « Analysis of Sets of Two-Way Contingency Tables Using Association Models », *Journal of the American Statistical Association*, 84, 405, p. 142-151.
- BECK U., BECK-GERNSHEIM E., 1995, *The Normal Chaos of Love*, Cambridge, Polity Press.
- BERENT J., 1954, « Social Mobility and Marriage: A Study of Trends in England and Wales », dans GLASS D.V. (dir.), *Social Mobility in Britain*, London, Routledge, p. 321-338.
- BERGSTRÖM M., 2011, « La toile des sites de rencontres en France », *Réseaux*, 166, 2, p. 225-260.
- BERGSTRÖM M., 2014, *Au bonheur des rencontres. Sexualité, classe et rapport de genre dans la production et l'usage des sites de rencontres en France*, Thèse de doctorat, Paris, Institut d'études politiques de Paris.
- BERNARDI F., 2001, « Is it a Timing or a Probability Effect? Four Simulations and an Application of Transition Rate Models to the Analysis of Unemployment Exit », *Quality and Quantity*, 35, 3, p. 231-252.
- BERNARDI F., 2003, « Who Marries Whom in Italy? », dans BLOSSFELD H.-P., TIMM A. (dirs.), *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht, Kluwer Academic Publishers, p. 113-139.
- BERNARDI L., OPPO A., 2011, « Couple Formation as a Transition Between Families », dans JALLINOJA R., WIDMER É.D. (dirs.), *Families and Kinship in Contemporary Europe. Rules and Practices of Relatedness*, Basingstoke, Palgrave Macmillan (Studies in Family and Intimate Life), p. 95-111.
- BERTAUX-WIAME I., 2004, « Devenir indépendant, une affaire de couple », *Cahiers du Genre*, 37, 2, p. 13-40.
- BÈS M.-P., 2013, « Les relations entre anciens élèves ingénieurs: réseau personnel ou capital social? », *Socio-logos*, 8.

- BESSIÈRE C., 2014, « Female and male domestic partners in wine-grape farms (Cognac, France): conjugal asymmetry and gender discrimination in family businesses », *The History of the Family*, 19, 3, p. 341-357.
- BEYERSMANN J., SCHUMACHER M., 2008, « Time-dependent covariates in the proportional subdistribution hazards model for competing risks », *Biostatistics*, 9, 4, p. 765-776.
- BIDART C., 2008, « Dynamiques des réseaux personnels et processus de socialisation : évolutions et influences des entourages lors des transitions vers la vie adulte », *Revue française de sociologie*, 49, 3, p. 559-583.
- BIDART C., LAVENU D., 2005, « Evolutions of personal networks and life events », *Social Networks*, 27, 4, p. 359-376.
- BIDART C., PELLISSIER A., 2002, « Copains d'école, copains de travail : Évolution des modes de sociabilité d'une cohorte de jeunes », *Réseaux*, 115, 5, p. 17-49.
- BIDOU-ZACHARIASEN C., 2000, « À propos de la "service class" : les classes moyennes dans la sociologie britannique », *Revue française de sociologie*, 41, 4, p. 777-796.
- BIHR A., PFEFFERKORN R., 2002, *Hommes, femmes, quelle égalité ? École, travail, couple, espace public*, Paris, L'Atelier.
- BILLARI F., APARICIO DIAZ B., FENT T., PRSKAWETZ A., 2007, « The "Wedding-Ring": An agent-based marriage model based on social interaction », *Demographic Research*, 17, 3, p. 59-82.
- BILLARI F.C., 2001, « A Log-Logistic Regression Model for a Transition Rate with a Starting Threshold », *Population Studies*, 55, 1, p. 15-24.
- BILLARI F.C., LIEFBROER A.C., 2010, « Towards a new pattern of transition to adulthood? », *Advances in Life Course Research*, 15, 2-3, p. 59-75.
- BILLARI F.C., PHILIPPOV D., BAIZÁN P., 2001, « Leaving Home in Europe: The Experience of Cohorts Born Around 1960 », *International Journal of Population Geography*, 7, 5, p. 339-356.
- BIRKELUND G.E., HELDAL J., 2003, « Who Marries Whom? Educational Homogamy in Norway », *Demographic Research*, 8, 1, p. 1-30.
- BISHOP Y.M.M., FIENBERG S.E., HOLLAND P.W., [1975] 2007, *Discrete Multivariate Analysis: Theory and Practice*, New York, Springer.
- BLACKWELL D.L., LICHTER D.T., 2000, « Mate Selection Among Married and Cohabiting Couples », *Journal of Family Issues*, 21, 3, p. 275-302.
- BLANPAIN N., 2011, « L'espérance de vie s'accroît, les inégalités sociales face à la mort demeurent », *Insee Première*, 1372.
- BLAU P.M., 1994, *Structural Contexts of Opportunities*, Chicago, University of Chicago Press.
- BLAU P.M., DUNCAN O.D., 1967, *The American Occupational Structure*, New York, Wiley.
- BLAU P.M., SCHWARTZ J.E., 1984, *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations*, Orlando, Academic Press.
- BLOSSFELD, H.-P. (dir.), 1995, *The New Role of Women: Family Formation in Modern Societies*, Boulder, Westview Press.

- BLOSSFELD H.-P., 2009, « Educational Assortative Marriage in Comparative Perspective », *Annual Review of Sociology*, 35, p. 513-530.
- BLOSSFELD H.-P., ROHWER G., [1995] 2002, *Techniques of Event History Modeling: New Approaches to Causal Analysis*, 2^e édition, Mahwah, Erlbaum.
- BLOSSFELD H.-P., TIMM A., 2003a, « Educational Systems as Marriage Markets in Modern Societies: a Conceptual Framework », *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht, Kluwer Academic Publishers, p. 1-18.
- BLOSSFELD, H.-P., TIMM, A. (dirs.), 2003b, *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht, Kluwer Academic Publishers.
- BLOSSFELD H.-P., TIMM A., 2003c, « Who Marries Whom in West Germany? », *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht, Kluwer Academic Publishers, p. 19-35.
- BLOSSFELD H.-P., TIMM A., DASKO F., 1997, « The Educational System as a Marriage Market. A Longitudinal Analysis of Marriage in the Life Course », Sonderforschungsbereich 186 Working Paper, 46, Bremen, Universität Bremen.
- BOCQUIER P., 2006, « Les effets peuvent-ils précéder les causes? Traitement des intentions et des anticipations », dans ANTOINE P., LELIÈVRE É. (dirs.), *États flous et trajectoires complexes. Observation, modélisation, interprétation*, Paris, Ined (Méthodes et savoirs), p. 239-260.
- BOLOGNE J.-C., 2004, *Histoire du célibat et des célibataires*, Paris, Fayard.
- BONVALET C., CLÉMENT C., OGG J., 2011, *Réinventer la famille: l'histoire des baby-boomers*, Paris, Presses Universitaires de France (Le lien social).
- BONVALET C., LELIÈVRE É., 1995, « Du concept de ménage à celui d'entourage: une redéfinition de l'espace familial », *Sociologie et sociétés*, 27, 2, p. 177-190.
- BONVALET C., LELIÈVRE É., 2012, *De la famille à l'entourage: l'enquête Biographies et entourage*, Paris, Ined (Grandes enquêtes).
- BOUCHET-VALAT M., 2011, *L'homogamie d'éducation et de statut social dans la France contemporaine*, Master 2, Paris, EHESS.
- BOUCHET-VALAT M., 2014a, « Les évolutions de l'homogamie de diplôme, de classe et d'origine sociales en France (1969-2011): ouverture d'ensemble, repli des élites », *Revue française de sociologie*, 55, 3, p. 459-505.
- BOUCHET-VALAT M., 2014b, « Socioeconomic Homogamy in the European Union: Assessment of ESeG Level 1 Prototypes and Indications Regarding Level 2 Categories Using Data from the European Labour Force Survey », dans MERON M., AMAR M. (dirs.), *Final Report of the European Statistical System Network on the Harmonisation and Implementation of a European Socio-Economic Classification: European Socio-Economic Groups (ESeG)*, Kirchberg, Eurostat.
- BOUCHET-VALAT M., 2014c, « Socioeconomic Homogamy in the European Union Seen Through the ESeG Classification », dans MERON M., AMAR M. (dirs.), *European Socio-Economic Groups (ESeG): first uses*, Kirchberg, Eurostat.

- BOUCHET-VALAT M., 2015a, *logmult: Log-multiplicative models, including association models*, R package, version 0.6.2.
- BOUCHET-VALAT M., 2015b, « L'analyse statistique des tables de contingence carrées - L'homogamie socioprofessionnelle en France - I. L'analyse des correspondances », *Bulletin de Méthodologie Sociologique*, 125, p. 65-88.
- BOUCHET-VALAT M., 2015c, « L'analyse statistique des tables de contingence carrées - L'homogamie socioprofessionnelle en France II. L'apport des modèles d'association », *Bulletin de Méthodologie Sociologique*, 126, p. 5-27.
- BOUFFARTIGUE P., 1999, « La crise d'un salariat de confiance. Les cadres déstabilisés », *Actuel Marx*, 26, p. 53-68.
- BOUFFARTIGUE P., 2001a, *Les cadres: fin d'une figure sociale*, Paris, La Dispute.
- BOUFFARTIGUE P., 2001b, « Les transformations d'un salariat de confiance », dans BOUFFARTIGUE P. (dir.), *Cadres: la grande rupture*, Paris, La Découverte (Recherches), p. 35-49.
- BOUFFARTIGUE P., GADEA, C., POCHIC, S. (dirs.), 2011, *Cadres, classes moyennes: vers l'éclatement ?*, Paris, Armand Colin (Recherches).
- BOURDIEU P., 1962, « Célibat et condition paysanne », *Études rurales*, p. 32-135.
- BOURDIEU P., 1966, « Condition de classe et position de classe », *Archives européennes de sociologie*, 7, 2, p. 201-223.
- BOURDIEU P., 1972, « Les stratégies matrimoniales dans le système de reproduction », *Annales. Histoire, Sciences Sociales*, 27, 4/5, p. 1105-1127.
- BOURDIEU P., 1974, « Avenir de classe et causalité du probable », *Revue française de sociologie*, 15, 1, p. 3-42.
- BOURDIEU P., 1979, *La Distinction. Critique sociale du jugement*, Le sens commun, Paris, Minuit.
- BOURDIEU P., 1980, *Le Sens pratique*, Paris, Minuit (Le sens commun).
- BOURDIEU P., 1984a, « Espace social et genèse des "classes" », *Actes de la recherche en sciences sociales*, 52, 1, p. 3-14.
- BOURDIEU P., 1984b, « La "jeunesse" n'est qu'un mot », *Questions de sociologie*, Paris, Minuit, p. 143-154.
- BOURDIEU P., 1989, *La Noblesse d'État: Grandes écoles et esprit de corps*, Paris, Minuit (Le sens commun).
- BOURDIEU P., 1998, *La domination masculine*, Paris, Seuil (Liber).
- BOURGUIGNON F., 1979, « Decomposable income inequality measures », *Econometrica*, 47, 4, p. 901-920.
- BOX G.E.P., DRAPER N.R., 1987, *Empirical Model-Building and Response Surfaces*, New York, Wiley.
- BOZON M., 1988, « Le mariage en moins », *Société française*, 26, p. 9-19.

- BOZON M., 1990a, « Les femmes et l'écart d'âge entre conjoints. Une domination consentie : II. Modes d'entrée dans la vie adulte et représentations du conjoint », *Population*, 45, 3, p. 565-602.
- BOZON M., 1990b, « Les femmes et l'écart d'âge entre conjoints. Une domination consentie : I. Types d'union et attentes en matière d'écart d'âge », *Population*, 45, 2, p. 327-360.
- BOZON M., 1991a, « Apparence physique et choix du conjoint », dans HIBERT T., ROUSSEL L. (dirs.), *La nuptialité, évolution récente en France et dans les pays développés. Actes du IX^e Colloque national de démographie*, Ined, p. 91-110.
- BOZON M., 1991b, « Mariage et mobilité sociale en France », *European Journal of Population*, 7, 2, p. 171-190.
- BOZON M., HÉRAN F., 1987a, « L'aire de recrutement du conjoint », *Données sociales 1987*, Paris, Insee, p. 338-347.
- BOZON M., HÉRAN F., 1987b, « La découverte du conjoint: I. Évolution et morphologie des scènes de rencontre », *Population*, 42, 6, p. 943-985.
- BOZON M., HÉRAN F., 1988, « La découverte du conjoint: II. Les scènes de rencontre dans l'espace social », *Population*, 43, 1, p. 121-150.
- BOZON M., HÉRAN F., 2006, *La formation du couple: Textes essentiels pour la sociologie de la famille*, Paris, La Découverte (Grands Repères Classiques).
- BOZON M., RAULT W., 2012, « De la sexualité au couple », *Population*, 67, 3, p. 453-490.
- BRAINERD E., 2000, « Women in Transition: Changes in Gender Wage Differentials in Eastern Europe and the Former Soviet Union », *Industrial & Labor Relations Review*, 54, 1, p. 138-162.
- BRANCA P., 1977, « Review: The Making of the Modern Family, by Edward Shorter, and The Family in History, Edited by Charles E. Rosenberg », *The Journal of Modern History*, 49, 1, p. 118-120.
- BRAUNS H., STEINMANN S., 1999, « Educational Reform in France, West-Germany, the United Kingdom and Hungary: Updating the CASMIN Educational Classification », *ZUMA Nachrichten*, 23, 44, p. 7-44.
- BRÉCHON P., 2013, « L'individualisation des sociétés européennes », *Futuribles*, 395, p. 119-136.
- BRÉCHON, P., GALLAND, O. (dirs.), 2010, *L'individualisation des valeurs*, Paris, Armand Colin.
- BREEN, R. (dir.), 2004, *Social Mobility in Europe*, Oxford, Oxford University Press.
- BREEN R., ANDERSEN S.H., 2012, « Educational Assortative Mating and Income Inequality in Denmark », *Demography*, 49, 3, p. 867-887.
- BREEN R., JONSSON J.O., 2000, « Analyzing Educational Careers: A Multinomial Transition Model », *American Sociological Review*, 65, 5, p. 754-772.
- BREEN R., LUIJKX R., MÜLLER W., POLLAK R., 2009, « Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries », *American Journal of Sociology*, 114, 5, p. 1475-1521.

- BREEN R., SALAZAR L., 2010, « Has Increased Women's Educational Attainment Led to Greater Earnings Inequality in the United Kingdom? A Multivariate Decomposition Analysis », *European Sociological Review*, 26, 2, p. 143-157.
- BREEN R., SALAZAR L., 2011, « Educational Assortative Mating and Earnings Inequality in the United States », *American Journal of Sociology*, 117, 3, p. 808-843.
- BREUIL-GENIER P., 2001, « Les variations des revenus individuels entre périodes de chômage et d'emploi », *Économie et statistique*, 348, 1, p. 61-79.
- BRÜDERL J., DIEKMANN A., 1995, « The Log-Logistic Rate Model », *Sociological Methods & Research*, 24, 2, p. 158-186.
- BRÜDERL J., DIEKMANN A., 1997, « Education and Marriage: A Comparative Study », Working Paper.
- BUFFETEAU S., AFSA C., 2006, « L'activité féminine en France: quelles évolutions récentes, quelles tendances pour l'avenir? », *Économie et statistique*, 398, 1, p. 85-97.
- BUISSON G., DAGUET F., 2012, « Qui vit seul dans son logement? Qui vit en couple? », *Insee Première*, 1392.
- BUISSON G., LAPINTE A., 2013, « Le couple dans tous ses états. Non-cohabitation, conjoints de même sexe, Pacs... », *Insee Première*, 1435.
- BUKODI E., RÓBERT P., ALTORJAI S., 2008, « The Hungarian educational system and the implementation of the ISCED-97 », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 200-215.
- BURNHAM K.P., ANDERSON D.R., 2004, « Multimodel Inference: Understanding AIC and BIC in Model Selection », *Sociological Methods & Research*, 33, 2, p. 261-304.
- BURNHAM K.P., ANDERSON D.R., 2014, « P values are only an index to evidence: 20th- vs. 21st-century statistical science », *Ecology*, 95, 3, p. 627-630.
- BURNOD G., CHENU A., 2001, « Employés qualifiés et non qualifiés: une proposition d'aménagement de la nomenclature des catégories socioprofessionnelles », *Travail et emploi*, 86, p. 87-105.
- BURTLESS G., 1999, « Effects of growing wage disparities and changing family composition on the U.S. income distribution », *European Economic Review*, 43, 4-6, p. 853-865.
- BUSS D.M., SHACKELFORD T.K., KIRKPATRICK L.A., LARSEN R.J., 2001, « A Half Century of Mate Preferences: The Cultural Evolution of Values », *Journal of Marriage and Family*, 63, 2, p. 491-503.
- CABANNES P.-Y., 2014, « Trois décennies d'évolutions du marché du travail », *Trente ans de vie économique et sociale*, Paris, Insee (Insee Références), p. 55-67.
- CACOUAULT M., 1984, « Diplôme et célibat. Les femmes professeurs de lycée entre les deux guerres », dans FARGE A., KLAPISCH-ZUBER C. (dirs.), *Madame ou Mademoiselle?*, Paris, Montalba, p. 177-203.
- CAILLE J.-P., LEMAIRE S., VROLANT M.-C., 2002, « Filles et garçons face à l'orientation », *Note d'information de la Direction de la programmation et du développement du Ministère de l'Éducation nationale*, 12, p. 1-6.

- CANCIAN M., REED D., 1998, « Assessing the Effects of Wives' Earnings on Family Income Inequality », *Review of Economics and Statistics*, 80, 1, p. 73-79.
- CANCIAN M., REED D., 1999, « The impact of wives' earnings on income inequality: Issues and estimates », *Demography*, 36, 2, p. 173-184.
- CASSAN F., MAZUY M., CLANCHÉ F., 2005, « Refaire sa vie de couple est plus fréquent pour les hommes », dans LEFÈVRE C., FILHON A. (dirs.), *Histoires de familles, histoires familiales : les résultats de l'enquête famille de 1999*, Paris, Ined (Cahiers), p. 223-234.
- CASTEL R., [1995] 1999, *Les métamorphoses de la question sociale. Une chronique du salariat*, Paris, Gallimard (Folio essais).
- CASTEL R., 1999, « Pourquoi la classe ouvrière a-t-elle perdu la partie ? », *Actuel Marx*, 26, p. 15-24.
- CASTEL R., 2009, *La montée des incertitudes. Travail, protections, statut de l'individu*, Paris, Seuil (La couleur des idées).
- CAUSSINUS H., 1965, « Contribution à l'analyse statistique des tableaux de corrélation », *Annales de la faculté des sciences de Toulouse*, 29, p. 77-183.
- CÉZARD M., 1986, « Le chômage et son halo », *Économie et statistique*, 193, 1, p. 77-82.
- CHAMPAGNE C., PAILHÉ A., SOLAZ A., 2014, « 25 ans de participation des hommes et des femmes au travail domestique : quels facteurs d'évolutions ? », *Documents de travail*, 203, Paris, Ined.
- CHAN T.W., HALPIN B., 2003, « Who Marries Whom in Great Britain? », dans BLOSSFELD H.-P., TIMM A. (dirs.), *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht, Kluwer Academic Publishers, p. 171-194.
- CHARNOZ P., COUDIN É., GAINI M., 2013, « Une diminution des disparités salariales en France entre 1967 et 2009 », *Emploi et salaires 2013*, Insee, Paris (Insee Références), p. 75-86.
- CHAUVEL L., 1998, « La seconde explosion scolaire : diffusion des diplômes, structure sociale et valeur des titres », *Revue de l'OFCE*, 66, 1, p. 5-36.
- CHAUVEL L., 2001, « Le retour des classes sociales ? », *Revue de l'OFCE*, 4, 79, p. 315-359.
- CHAUVEL L., 2006, « La déstabilisation du système de positions sociales », dans LAGRANGE H. (dir.), *L'épreuve des inégalités*, Paris, Presses Universitaires de France (Le lien social), p. 91-112.
- CHENU A., 1990, *L'archipel des employés*, Paris, Insee.
- CHRISTENSEN R., 1994, *Log-Linear Models*, New York, Springer.
- CLAIR I., 2008, *Les jeunes et l'amour dans les cités*, Paris, Armand Colin.
- CLOGG C.C., 1982, « Some Models for the Analysis of Association in Multiway Cross-Classifications Having Ordered Categories », *Journal of the American Statistical Association*, 77, 380, p. 803-815.
- CLOGG C.C., SHIHADDEH E.S., 1994, *Statistical Models for Ordinal Variables*, Sage (Advanced Quantitative Techniques in the Social Sciences).
- COALE A.J., MCNEIL D.R., 1972, « The Distribution by Age of the Frequency of First Marriage in a Female Cohort », *Journal of the American Statistical Association*, 67, 340, p. 743-749.

- COBALTI A., 1989, « A relative mobility table », *Quality and Quantity*, 23, 2, p. 205-220.
- COLES M.G., FRANCESCONI M., 2011, « On the Emergence of Toyboys: the Timing of Marriage With Aging and Uncertain Careers », *International Economic Review*, 52, 3, p. 825-853.
- COMMAILLE J., 1974, « Divorce, milieu social et situation de la femme », *Économie et statistique*, 53, p. 3-21.
- CONSTANTINE A.G., GOWER J.C., 1978, « Graphical Representation of Asymmetric Matrices », *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 27, 3, p. 297-304.
- COONEY T.M., HOGAN D.P., 1991, « Marriage in an Institutionalized Life Course: First Marriage among American Men in the Twentieth Century », *Journal of Marriage and Family*, 53, 1, p. 178-190.
- CORTESE G., ANDERSEN P.K., 2010, « Competing Risks and Time-Dependent Covariates », *Biometrical Journal*, 52, 1, p. 138-158.
- COULANGEON P., MENGER P.-M., ROHARIK I., 2002, « Les loisirs des actifs: un reflet de la stratification sociale », *Économie et statistique*, 352, 1, p. 39-55.
- COUPPIÉ T., EPIPHANE D., 2006, « La ségrégation des hommes et des femmes dans les métiers: entre héritage scolaire et construction sur le marché du travail », *Formation emploi. Revue française de sciences sociales*, 93, p. 11-27.
- COURGEAU D., 1987, « Constitution de la famille et urbanisation », *Population*, 42, 1, p. 57-81.
- COURGEAU D., LELIÈVRE É., 1986, « Nuptialité et agriculture », *Population*, 41, 2, p. 303-326.
- COURGEAU D., LELIÈVRE É., 1988, « Estimation of transition rates in dynamic household models », dans KEILMAN N., KUIJSTEN A., VOSSEN A. (dirs.), *Modelling Household Formation and Dissolution*, Oxford, Clarendon Press, p. 160-176.
- COURGEAU D., LELIÈVRE É., 1989, *Analyse démographique des biographies*, Paris, Ined (Manuels).
- COURSON J.-P., 1982, « Les ménages n'auront plus de chef », *Économie et statistique*, 149, p. 47-55.
- COX D.R., 1972, « Regression Models and Life-Tables », *Journal of the Royal Statistical Society. Series B (Methodological)*, 34, 2, p. 187-220.
- DAGUET F., NIEL X., 2010, « La proportion de jeunes en couple se stabilise », *Insee Première*, 1281.
- DAUMAS M., 2004, *Le mariage amoureux: Histoire du lien conjugal sous l'Ancien Régime*, Paris, Armand Colin.
- DAVIDOV E., SCHMIDT P., SCHWARTZ S.H., 2008, « Bringing Values Back In. The Adequacy of the European Social Survey to Measure Values in 20 Countries », *Public Opinion Quarterly*, 72, 3, p. 420-445.
- DÉCHAUX J.-H., 1999, « Les deux faces de l'individualisme familial », *Les Cahiers de la sécurité intérieure*, 35, p. 11-32.
- DELPHY C., 1977, « Les femmes dans les études de stratification », dans MICHEL A. (dir.), *Femmes, sexisme et sociétés*, Paris, Presses Universitaires de France (Sociologie d'aujourd'hui), p. 25-38.

- DESPLANQUES G., 1987, « Cycle de vie et milieu social », Collections, *Série D*, 117, Paris, Insee.
- DESROSIÈRES A., 1978, « Marché matrimonial et structure des classes sociales », *Actes de la recherche en sciences sociales*, 20, 1, p. 97-107.
- DESROSIÈRES A., 1995, « Classer et mesurer: les deux faces de l'argument statistique », *Réseaux*, 13, 71, p. 11-29.
- DESROSIÈRES A., 2001, « Entre réalisme métrologique et conventions d'équivalence: les ambiguïtés de la sociologie quantitative », *Genèses*, 43, 2, p. 112-127.
- DEVEREUX P.J., 2004, « Changes in relative wages and family labor supply », *Journal of Human Resources*, 39, 3, p. 698-722.
- DEVILLE J.-C., 1981, « De l'enfance à la constitution d'une famille », *Données sociales 1981*, Paris, Insee, p. 17-43.
- DEWERPE A., 1996, « La "stratégie" chez Pierre Bourdieu. Note de lecture », *Enquête*, 3, p. 191-208.
- DIEKMANN A., 1989, « Diffusion and survival models for the process of entry into marriage », *The Journal of Mathematical Sociology*, 14, 1, p. 31-44.
- DIMAGGIO P., MOHR J., 1985, « Cultural Capital, Educational Attainment, and Marital Selection », *The American Journal of Sociology*, 90, 6, p. 1231-1261.
- DOMAŃSKI H., PRZYBYSZ D., 2007, « Educational Homogamy in 22 European Countries », *European Societies*, 9, 4, p. 495-526.
- DUBET F., 2001, « Les inégalités multipliées ou les épreuves de l'égalité », dans MERCURE D. (dir.), *Une société-monde ?*, De Boeck Supérieur (Ouvertures sociologiques), p. 93-114.
- DUFOUR-KIPPELEN S., DORMONT B., 2000, « Insertion professionnelle et autonomie résidentielle: le cas des jeunes peu diplômés », *Économie et statistique*, 337, p. 97-120.
- DUNCAN O.D., 1979, « How Destination Depends on Origin in the Occupational Mobility Table », *American Journal of Sociology*, 84, 4, p. 793-803.
- DURKHEIM É., [1892] 1975, « La famille conjugale », *Textes*, 3. Fonctions sociales et institutions, Paris, Minuit (Le sens commun), p. 35-49.
- DURKHEIM É., [1893] 2013, *De la division du travail social*, 8^e édition, Paris, Presses Universitaires de France (Quadrige).
- DURU-BELLAT M., 1990, *L'école des filles. Quelle formation pour quels rôles sociaux ?*, Paris, L'Harmattan.
- DURU-BELLAT M., 2004, « Ecole de garçons et école de filles... », *Diversité - Ville, école, intégration, enjeux*, 138, p. 65-72.
- EBBINGHAUS B., 2005, « When Less is More. Selection Problems in Large-N and Small-N Cross-National Comparisons », *International Sociology*, 20, 2, p. 133-152.
- ECKERT H., 2010, « "Précarité" dites-vous ? », *SociologieS*, disponible sur <https://sociologies.revues.org/3285>, mis en ligne le 27 septembre 2010.
- ECKERT H., MORA V., 2008, « Formes temporelles de l'incertitude et sécurisation des trajectoires dans l'insertion professionnelle des jeunes », *Travail et Emploi*, 113, p. 31-57.

- EIKA L., MOGSTAD M., ZAFAR B., 2014, « Educational assortative mating and household income inequality », Staff Report, 682, New York, Federal Reserve Bank of New York.
- ERIKSON R., GOLDTHORPE J.H., 1987, « Commonality and variation in social fluidity in industrial nations. Part I: A model for evaluating the 'FJH hypothesis' », *European Sociological Review*, 3, 1, p. 54-77.
- ERIKSON R., GOLDTHORPE J.H., 1992, *The Constant Flux: A Study of Class Mobility in Industrial Societies*, Oxford, Clarendon Press.
- ERIKSON R., GOLDTHORPE J.H., PORTOCARERO L., 1979, « Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden », *The British Journal of Sociology*, 30, 4, p. 415-441.
- ERLICH V., 1998, *Les nouveaux étudiants. Un groupe social en mutation*, Paris, Armand Colin.
- EROLA J., HÄRKÖNEN J., 2013, « Educational homogamy and entry into first marriage: analyzing discrete-time event-history data with conditional logit models », RC28 Spring Meeting, Trento, mai 2013.
- EROLA J., HÄRKÖNEN J., DRONKERS J., 2012, « More Careful or Less Marriageable? Parental Divorce, Spouse Selection and Entry into Marriage », *Social Forces*, 90, 4, p. 1323-1345.
- ESPING-ANDERSEN G., 2007, « Multiple Regression in Small-N Comparisons », *Comparative Social Research*, 24, p. 335-342.
- ESTEVE A., CORTINA C., 2006, « Changes in educational assortative mating in contemporary Spain », *Demographic Research*, 14, 17, p. 405-428.
- ESTEVE A., GARCÍA-ROMÁN J., PERMANYER I., 2011, « The reversal of the gender gap in education and its impact on union formation: the end of hypergamy », Population Association of America Annual Meeting, Washington, DC, 2 avril 2011.
- ESTEVE A., GARCÍA-ROMÁN J., PERMANYER I., 2012, « The Gender-Gap Reversal in Education and Its Effect on Union Formation: The End of Hypergamy? », *Population and Development Review*, 38, 3, p. 535-546.
- ESTEVE A., MCCAA R., 2007, « Homogamia educacional en México y Brasil, 1970-2000: Pautas y tendencias », *Latin American Research Review*, 42, 2, p. 56-85.
- ESTEVE A., MCCAA R., LÓPEZ L.Á., 2013, « The Educational Homogamy Gap Between Married and Cohabiting Couples in Latin America », *Population Research and Policy Review*, 32, 1, p. 81-102.
- EURIAT M., THÉLOT C., 1995, « Le recrutement social de l'élite scolaire en France. Évolution des inégalités de 1950 à 1990 », *Revue française de sociologie*, 36, 3, p. 403-438.
- EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2012, « The situation of Roma in 11 EU Member States ».
- FALGUEROLLES A. DE, HEIJDEN P.G.M. VAN DER, 2002, « Reduced rank quasi-symmetry and quasi-skew symmetry: a generalized bi-linear model approach », *Annales de la faculté des sciences de Toulouse – Mathématiques*, 11, 4, p. 507-524.
- FARBER B., 1977, « Review: The Making of the Modern Family, by Edward Shorter, and The Family in History, Edited by Charles E. Rosenberg », *Contemporary Sociology*, 6, 2, p. 167-169.

- FEATHERMAN D.L., JONES F.L., HAUSER R.M., 1975, « Assumptions of Social Mobility Research in the US: The Case of Occupational Status », *Social Science Research*, 4, 4, p. 329-360.
- FINE J.P., GRAY R.J., 1999, « A Proportional Hazards Model for the Subdistribution of a Competing Risk », *Journal of the American Statistical Association*, 94, 446, p. 496-509.
- FIRTH D., KUHA J., 1999, « Comments on "A Critique of the Bayesian Information Criterion for Model Selection" », *Sociological Methods & Research*, 27, 3, p. 398-402.
- FIRTH D., MENEZES R.X.D., 2004, « Quasi-Variations », *Biometrika*, 91, 1, p. 65-80.
- FLANDRIN J.-L., 1975, *Les Amours paysannes. Amour et sexualité dans les campagnes de l'ancienne France (XVI^e-XIX^e siècle)*, Paris, Gallimard (Archives).
- FLANQUART H., 1999, « Un désert matrimonial. Le célibat des jeunes femmes d'origine maghrébine en France », *Terrain*, 33, p. 127-144.
- FORSÉ M., CHAUVEL L., 1995, « L'évolution de l'homogamie en France. Une méthode pour comparer les diagonalités de plusieurs tables », *Revue française de sociologie*, 36, 1, p. 123-142.
- FOUQUET A., 2005, « Les femmes chefs d'entreprise: le cas français », *Travail, genre et sociétés*, 13, p. 31-50.
- FOX J., 2000, *Nonparametric Simple Regression: Smoothing Scatterplots*, Thousand Oaks, Sage (Quantitative Applications in the Social Sciences).
- FRÉMEAUX N., 2014, « Le rôle de l'héritage et du revenu du travail dans les choix matrimoniaux », *Population*, 69, 4, p. 551-587.
- FRIENDLY M., 1994, « Mosaic displays for multi-way contingency tables », *Journal of the American Statistical Association*, 89, 425, p. 190-200.
- FUNES LEAL V., 2015, *Efectos del Emparejamiento Selectivo en la Distribución del Ingreso en América Latina*, Thesis de Maestría en Economía, La Plata, Universidad Nacional de La Plata.
- GADEA C., MARRY C., 2000, « Les pères qui gagnent », *Travail, genre et sociétés*, 3, p. 109-135.
- GALLAND O., [1991] 2011, *Sociologie de la jeunesse*, 5^e édition, Paris, Armand Colin.
- GALLAND O., 1995, « Une entrée de plus en plus tardive dans la vie adulte », *Économie et statistique*, 283, p. 33-52.
- GALLAND O., 2000, « Une polarisation de la jeunesse française », *Revue de l'OFCE*, 72, 1, p. 221-228.
- GANZEBOOM H.B., TREIMAN D.J., ULTEE W.C., 1991, « Comparative Intergenerational Stratification Research: Three Generations and Beyond », *Annual Review of Sociology*, p. 277-302.
- GASQUET C., 2003, « Les jeunes « sans qualification »: un groupe hétérogène, des parcours d'insertion divers », *Bref Céreq*, 202.
- GELMAN A., RUBIN D.B., 1995, « Avoiding model selection in Bayesian social research », *Sociological Methodology*, 25, p. 165-174.

- GELMAN A., RUBIN D.B., 1999, « Evaluating and Using Statistical Methods in the Social Sciences. A Discussion of “A Critique of the Bayesian Information Criterion for Model Selection” », *Sociological Methods & Research*, 27, 3, p. 403-410.
- GIDDENS A., 2004, *La Transformation de l'intimité: Sexualité, amour et érotisme dans les sociétés modernes*, Editions du Rouergue.
- GILULA Z., HABERMAN S.J., 1986, « Canonical Analysis of Contingency Tables by Maximum Likelihood », *Journal of the American Statistical Association*, 81, 395, p. 780-788.
- GIRARD A., 1964, *Le Choix du conjoint: une enquête psycho-sociologique en France*, Paris, Presses Universitaires de France.
- GIRAUD C., RÉMY J., 2008, « Les choix des conjoints en agriculture », *Revue d'Études en Agriculture et Environnement*, 88, 3, p. 21-46.
- GOBLOT E., 1925, *La barrière et le niveau. Étude sociologique sur la bourgeoisie française moderne*, Paris, Alcan.
- GODECHOT O., 2012, « Is finance responsible for the rise in wage inequality in France? », *Socio-Economic Review*, 10, 3, p. 447-470.
- GODECHOT O., 2013, « Financiarisation et fractures socio-spatiales », *L'Année sociologique*, 63, 1, p. 17-50.
- GOFFMAN E., [1977] 2002, *L'arrangement des sexes*, Paris, La Dispute.
- GOLDMAN N., WESTOFF C.F., HAMMERSLOUGH C., 1984, « Demography of the marriage market in the United States », *Population Index*, p. 5-25.
- GOLDSCHIEDER F.K., WAITE L.J., 1986, « Sex Differences in the Entry Into Marriage », *American Journal of Sociology*, 92, 1, p. 91-109.
- GOLDSTEIN J.R., KENNEY C.T., 2001, « Marriage Delayed or Marriage Forgone? New Cohort Forecasts of First Marriage for U.S. Women », *American Sociological Review*, 66, 4, p. 506-519.
- GOLDTHORPE J.H., 1997, « Current Issues in Comparative Macrosociology: A Debate On Methodological Issues », *Comparative Social Research*, 16, p. 1-26.
- GOLLAC M., 1989, « Les dimensions de l'organisation du travail », *Économie et statistique*, 224, p. 27-44.
- GOODE W.J., 1959, « The Theoretical Importance of Love », *American Sociological Review*, 24, 1, p. 38-47.
- GOODE W.J., [1964] 1982, *The Family*, 2^e édition, Englewood Cliffs (New Jersey), Prentice-Hall (Foundations of Modern Sociology).
- GOODMAN L.A., 1965, « On the Statistical Analysis of Mobility Tables », *American Journal of Sociology*, 70, 5, p. 564-585.
- GOODMAN L.A., 1969, « How to Ransack Social Mobility Tables and Other Kinds of Cross-Classification Tables », *American Journal of Sociology*, 75, 1, p. 1-40.
- GOODMAN L.A., 1979a, « Simple Models for the Analysis of Association in Cross-Classifications having Ordered Categories », *Journal of the American Statistical Association*, 74, 367, p. 537-552.

- GOODMAN L.A., 1979b, « Multiplicative Models for the Analysis of Occupational Mobility Tables and Other Kinds of Cross-Classification Tables », *American Journal of Sociology*, 84, 4, p. 804-819.
- GOODMAN L.A., 1981, « Association models and canonical correlation in the analysis of cross-classifications having ordered categories », *Journal of the American Statistical Association*, 76, 374, p. 320-334.
- GOODMAN L.A., 1985, « The Analysis of Cross-Classified Data Having Ordered and/or Unordered Categories: Association Models, Correlation Models, and Asymmetry Models for Contingency Tables With or Without Missing Entries », *The Annals of Statistics*, 13, 1, p. 10-69.
- GOODMAN L.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*, 54, 3, p. 243-270.
- GOODMAN L.A., 1991, « Measures, Models, and Graphical Displays in the Analysis of Cross-Classified Data », *Journal of the American Statistical Association*, 86, 416, p. 1085-1111.
- GOODMAN L.A., 1996, « A Single General Method for the Analysis of Cross-Classified Data: Reconciliation and Synthesis of Some Methods of Pearson, Yule, and Fisher, and Also Some Methods of Correspondence Analysis and Association Analysis », *Journal of the American Statistical Association*, 91, 433, p. 408-428.
- GOODMAN L.A., HOUT M., 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*, 28, 1, p. 175-230.
- GOODMAN L.A., HOUT M., 2001, « Statistical Methods and Graphical Displays for Analyzing How the Association Between Two Qualitative Variables Differs Among Countries, Among Groups, Or Over Time. Part II: Some Exploratory Techniques, Simple Models, and Simple Examples », *Sociological Methodology*, 31, 1, p. 189-221.
- GOODY J., [1983] 1985, *L'Évolution de la famille et du mariage en Europe*, Paris, Armand Colin.
- GOODY J., 2001, *La famille en Europe*, Paris, Seuil (Faire l'Europe).
- GORDON M., 1977, « Review: The Making of the Modern Family, by Edward Shorter, and The Family in History, Edited by Charles E. Rosenberg », *Contemporary Sociology*, 6, 2, p. 169-171.
- GOUX D., 2003, « Une histoire de l'Enquête Emploi », *Économie et statistique*, 362, p. 41-57.
- GOUX D., CHARDON O., 2003, « La nouvelle définition européenne du chômage BIT », *Économie et statistique*, 362, p. 67-83.
- GOUX D., MAURIN É., 2003, « Who Marries Whom in France? », dans BLOSSFELD H.-P., TIMM A. (dirs.), *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht, Kluwer Academic Publishers, p. 57-78.
- GRAAF N.D. DE, SMEENK W., ULTEE W., TIMM A., 2003, « The When and Whom of First Marriage in The Netherlands », dans BLOSSFELD H.-P., TIMM A. (dirs.), *Who Marries*

- Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht, Kluwer Academic Publishers, p. 79-111.
- GRAVE B.S., SCHMIDT C.M., 2012, « The Dynamics of Assortative Mating in Germany », *Ruhr Economic Paper*, 346, Bochum.
- GREENACRE M., 2000, « Correspondence Analysis of Square Asymmetric Matrices », *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 49, 3, p. 297-310.
- GREENWOOD J., GUNER N., KOCHARKOV G., SANTOS C., 2014, « Marry Your Like: Assortative Mating and Income Inequality », *American Economic Review: Papers & Proceedings*, 104, 5, p. 348-353.
- GUIBERT-LANTOINE C. DE, LERIDON H., 1998, « La contraception en France. Un bilan après 30 ans de libéralisation », *Population*, 53, 4, p. 785-811.
- GUICHARD-CLAUDIC Y., TESTENOIRE A., TRANCART D., 2009, « Distances et proximités conjugales en situation d'homogamie et d'hétérogamie », dans PAILHÉ A., SOLAZ A. (dirs.), *Entre famille et travail: des arrangements de couple aux pratiques des employeurs*, Paris, La Découverte, p. 187-207.
- HABERMAN S.J., 1974, « Log-Linear Models for Frequency Tables with Ordered Classifications », *Biometrics*, 30, 4, p. 589-600.
- HALL J.R., 1954, « A Comparison of the Degree of Social Endogamy in England and Wales and the U.S.A. », dans GLASS D.V. (dir.), *Social Mobility in Britain*, London, Routledge, p. 344-346.
- HALPIN B., CHAN T.W., 2003, « Educational homogamy in Ireland and Britain: trends and patterns », *The British Journal of Sociology*, 54, 4, p. 473-495.
- HAMPLOVÁ D., 2009, « Educational Homogamy Among Married and Unmarried Couples in Europe: Does Context Matter? », *Journal of Family Issues*, 30, 1, p. 28-52.
- HAMPLOVÁ D., LE BOURDAIS C., 2008, « Educational Homogamy of Married and Unmarried Couples in English and French Canada », *Canadian Journal of Sociology*, 33, 4, p. 845-872.
- HARMENBERG K., 2014, « A Note: The Effect of Assortative Mating on Income Inequality ».
- HARRELL F.E., 2001, *Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis*, New York, Springer.
- HARRELL F.E., 2015, *rms: Regression Modeling Strategies*, R package, version 4.3-0.
- HARTIGAN J.A., KLEINER B., 1981, « Mosaics for contingency tables » dans EDDY W.F. (dir.), *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, p. 268-273.
- HEIJDEN P.G.M. VAN DER, MOOIJAART A., 1995, « Some New Log-Bilinear Models for the Analysis of Asymmetry in a Square Contingency Table », *Sociological Methods and Research*, 24, p. 7-29.
- HEIJDEN P.G.M. VAN DER, MOOIJAART A., TAKANE Y., 1994, « Correspondence analysis and contingency table models », dans GREENACRE M., BLASIUS J. (dirs.), *Correspondence Analysis in the Social Sciences. Recent Developments and Applications*, London-San Diego, Academic Press, p. 79-111.

- HENDRICKX J., 1998, « Religious and educational assortative marriage patterns in the Netherlands, 1940-1985 », *The Netherlands Journal of Social Sciences*, 34, 1, p. 5-22.
- HENZ U., JONSSON J.O., 2003, « Who Marries Whom in Sweden? », dans BLOSSFELD H.-P., TIMM A. (dirs.), *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*, Dordrecht, Kluwer Academic Publishers, p. 235-266.
- HÉRAN F., 1994, « Les mères persévèrent », *Insee Première*, 350.
- HERPIN N., 2003, « La taille des hommes: son incidence sur la vie en couple et la carrière professionnelle », *Économie et statistique*, 361, p. 71-90.
- HOU F., MYLES J., 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*, 33, 2, p. 337-366.
- HOUT M., BROOKS C., MANZA J., 1995, « The Democratic Class Struggle in the United States, 1948-1992 », *American Sociological Review*, 60, 6, p. 805-828.
- HRYSKO D., JUHN C., MCCUE K., 2014, « Trends in Earnings Inequality and Earnings Instability among US Couples: How Important Is Assortative Matching? », IZA Discussion Paper, 8729, Institute for the Study of Labor, Bonn.
- HUGRÉE C., PÉNISSAT É., SPIRE A., 2015, « Les différences entre salariés du public et du privé après le tournant managérial des États en Europe », *Revue française de sociologie*, 56, 1, p. 47-73.
- HYSLOP D.R., 2001, « Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings », *The American Economic Review*, 91, 4, p. 755-777.
- ICHOU M., VALLET L.-A., 2013, « Academic Achievement, Tracking Decisions, and Their Relative Contribution to Educational Inequalities », dans JACKSON M. (dir.), *Determined to Succeed? Performance versus Choice in Educational Attainment*, Stanford, Stanford University Press, p. 116-148.
- INSEE, 2013, « Marché du travail - Séries longues », *Insee Résultats*, 149.
- IVANČIČ A., 2008, « International comparability of the outcomes of the Slovenian education system – national vs international classification of education », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 176-199.
- JACKSON C., 2014, *flexsurv: Flexible parametric survival models*, R package, version 0.5.
- JAMIESON L., 1999, « Intimacy Transformed? A Critical Look at the “Pure Relationship” », *Sociology*, 33, 3, p. 477-494.
- JASPARD M., BROWN E., BATTAGLIOLA F., 1995, « Précarité d’emploi et itinéraires de transition à l’âge adulte », *Recherches et Prévisions*, 40, 1, p. 45-56.
- JEGOUZO G., 1972, « L’ampleur du célibat chez les agriculteurs », *Économie et statistique*, 34, p. 13-22.
- JEGOUZO G., 1991, « L’évolution après 1975 du célibat agricole masculin », *Population*, 46, 1, p. 41-61.

- JEGOUZO G., DAUCE P., 1969, « L'inertie professionnelle des chefs d'exploitation agricole », *Études rurales*, 36, 1, p. 37-65.
- JENKS G.F., 1967, « The Data Model Concept in Statistical Mapping », *International Yearbook of Cartography*, 7, 1, p. 186-190.
- JONES F.L., 1987, « Marriage Patterns and the Stratification System: Trends in Educational Homogamy since the 1930s », *Australian and New Zealand Journal of Sociology*, 23, 2, p. 185-198.
- JUHEM P., 1995, « Les relations amoureuses des lycéens », *Sociétés contemporaines*, 21, p. 29-42.
- KALMIJN M., 1991, « Status Homogamy in the United States », *The American Journal of Sociology*, 97, 2, p. 496-523.
- KALMIJN M., 1994, « Assortative Mating by Cultural and Economic Occupational Status », *The American Journal of Sociology*, 100, 2, p. 422-452.
- KALMIJN M., 1998, « Inter-marriage and Homogamy: Causes, Patterns, Trends », *Annual Review of Sociology*, 24, p. 395-421.
- KALMIJN M., FLAP H., 2001, « Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices », *Social Forces*, 79, 4, p. 1289-1312.
- KANEKO R., 2003, « Elaboration of the Coale-McNeil Nuptiality Model as the Generalized Log Gamma Distribution: A New Identity and Empirical Enhancements », *Demographic Research*, 9, 10, p. 221-262.
- KAPPI C., STRATOUDAKI H., 2004, « Educational system », dans CHARALAMBIS D., MARATOU-ALIPRANTI L., HADJIYANNI A. (dirs.), *Recent Social Trends in Greece, 1960-2000*, McGill-Queen's Press.
- KAROLY L.A., BURTLESS G., 1995, « Demographic change, rising earnings inequality, and the distribution of personal well-being, 1959-1989 », *Demography*, 32, 3, p. 379-405.
- KATRŇÁK T., 2008, « Educational Assortative Mating in the Czech Republic, Slovakia and Hungary between 1976 and 2003 », *Sociológia*, 40, 3, p. 1-24.
- KATRŇÁK T., FUČÍK P., LUIJKX R., 2012, « The relationship between educational homogamy and educational mobility in 29 European countries », *International Sociology*, 27, 4, p. 551-573.
- KATRŇÁK T., KREIDL M., FÓNADOVÁ L., 2006, « Trends in Educational Assortative Mating in Central Europe: the Czech Republic, Slovakia, Poland, and Hungary, 1988-2000 », *European Sociological Review*, 22, 3, p. 309-322.
- KAUFMANN J.-C., 1992, *La trame conjugale. Analyse du couple par son linge*, Paris, Nathan (Essais & Recherches).
- KAUFMANN J.-C., [1993] 2014, *Sociologie du couple*, 6^e édition, Paris, Presses Universitaires de France (Que sais-je?).
- KAUFMANN J.-C., 1999, *La femme seule et le prince charmant. Enquête sur la vie en solo*, Paris, Nathan (Essais & Recherches).

- KAUFMAN R.L., SCHERVISH P.G., 1987, « Variations on a Theme. More Uses of Odds Ratios to Interpret Log-Linear Parameters », *Sociological Methods & Research*, 16, 2, p. 218-255.
- KEELE L.J., 2008, *Semiparametric Regression for the Social Sciences*, New York, Wiley.
- KEELEY M.C., 1979, « An Analysis of the Age Pattern of First Marriage », *International Economic Review*, 20, 2, p. 527-544.
- KELLERHALS J., LANGUIN N., PERRIN J.-F., WIRTH G., 1985, « Statut social, projet familial et divorce: une analyse longitudinale des ruptures d'union dans une promotion de mariages », *Population*, 40, 6, p. 811-825.
- KELLERHALS J., PERRIN J.-F., STEINAUER-CRESSON G., VONÈCHE L., WIRTH G., 1982, *Mariages au quotidien: inégalités sociales, tensions culturelles et organisation familiale*, Lausanne, Favre.
- KEMPENEERS M., LELIÈVRE É., 1993, « Women's work in the EC: Five career profiles », *European Journal of Population*, 9, 1, p. 77-92.
- KERGOAT, D., GUICHARD-CLAUDIC, Y., VILBROD, A. (dirs.), 2008, *L'inversion du genre. Quand les métiers masculins se conjuguent au féminin... et réciproquement*, Rennes, Presses Universitaires de Rennes.
- KERR C., DUNLOP J.T., HARBISON F.H., MYERS C.A., 1960, *Industrialism Industrial Man*, Harvard, Harvard University Press.
- KIEFFER A., 2008, « Applying the ISCED-97 to France: some issues and propositions », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education*, Mannheim, MZES, p. 103-124.
- KOGAN, I., GEBEL, M., NOELKE, C. (dirs.), 2008, *Europe Enlarged: A Handbook of Education, Labour and Welfare Regimes in Central and Eastern Europe*, Bristol, Policy Press.
- KOGAN I., GEBEL M., NOELKE C., 2012, « Educational Systems and Inequalities in Educational Attainment in Central and Eastern European Countries », *Studies of Transition States and Societies*, 4, 1, p. 69-83.
- KOŠTA J., BEDNÁRIK R., 2008, « Slovakia », dans KOGAN I., GEBEL M., NOELKE C. (dirs.), *Europe Enlarged: A Handbook of Education, Labour and Welfare Regimes in Central and Eastern Europe*, Bristol, Policy Press, p. 323-352.
- KOSTOVA D., 2008, « The Bulgarian educational system and evaluation of the ISCED-97 implementation », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education*, Mannheim, MZES, p. 163-175.
- KUHA J., FIRTH D., 2011, « On the index of dissimilarity for lack of fit in loglinear and log-multiplicative models », *Computational Statistics & Data Analysis*, 55, 1, p. 375-388.
- LANDAIS C., 2007, « Les hauts revenus en France (1998-2006): une explosion des inégalités », Document de travail, Paris, École d'économie de Paris.
- LARRIMORE J., 2014, « Accounting for United States Household Income Inequality Trends: The Changing Importance of Household Structure and Male and Female Labor Earnings Inequality », *Review of Income and Wealth*, 60, 4, p. 683-701.

- LATOUCHE A., PORCHER R., CHEVRET S., 2005, « A Note on Including Time-dependent Covariate in Regression Model for Competing Risks Data », *Biometrical Journal*, 47, 6, p. 807-814.
- LE DONNÉ N., 2014, « La réforme de 1999 du système éducatif polonais. Effets sur les inégalités sociales de compétences scolaires », *Revue française de sociologie*, 55, 1, p. 127-162.
- LEEUWEN M.H.D. VAN, MAAS I., 2005, « Endogamy and Social Class in History: An Overview », *International Review of Social History*, 50, S13, p. 1-23.
- LEFEUVRE S., 2008, « Poursuite d'études et engagement conjugal en situation d'hypogamie », dans KERGOAT D., GUICHARD-CLAUDIC Y., VILBROD A. (dirs.), *L'inversion du genre. Quand les métiers masculins se conjuguent au féminin... et réciproquement*, Rennes, Presses Universitaires de Rennes, p. 83-95.
- LEFÈVRE, C., FILHON, A. (dirs.), 2005, *Histoires de familles, histoires familiales: les résultats de l'enquête famille de 1999*, Paris, INED (Cahiers).
- LELIÈVRE É., MAZUY M., 2005, « Déclarer ses enfants, déclarer ses conjoints », dans LEFÈVRE C., FILHON A. (dirs.), *Histoires de familles, histoires familiales: les résultats de l'enquête famille de 1999*, Paris, INED (Cahiers), p. 573-588.
- LELIÈVRE É., ROUBAUD F., TICHIT C., VIVIER G., 2006, « Données factuelles et perceptions, la place du flou dans l'observation et l'analyse », dans ANTOINE P., LELIÈVRE É. (dirs.), *États flous et trajectoires complexes. Observation, modélisation, interprétation*, Paris, Ined (Méthodes et savoirs), p. 29-60.
- LEMARCHANT C., 2001, « La construction du couple sous le regard des parents », *Agora débats/jeunesses*, 23, 1, p. 51-59.
- LENSKI G.E., 1954, « Status Crystallization: A Non-Vertical Dimension of Social Status », *American Sociological Review*, 19, 4, p. 405-413.
- LENSKI G., NOLAN P.D., 1984, « Trajectories of Development: A Test of Ecological-Evolutionary Theory », *Social Forces*, 63, 1, p. 1-23.
- LE PAPE M.-C., 2006, « Les ambivalences d'une double appartenance : hommes et femmes en milieux populaires », *Sociétés contemporaines*, 62, 2, p. 5-26.
- LERIDON H., CHARBIT Y., COLLOMB P., SARDON J.-P., TOULEMON L., 1987, *La seconde révolution contraceptive. La révolution des naissances en France de 1950 à 1985*, Paris, Ined (Travaux et documents).
- LETT D., 2000, *Famille et parenté dans l'Occident médiéval. V^e-XV^e siècle*, Paris, Hachette Éducation.
- LICHTER D.T., 1990, « Delayed Marriage, Marital Homogamy, and the Mate Selection Process among White Women », *Social Science Quarterly*, 71, 4, p. 802-811.
- LIEBETRAU A.M., 1983, *Measures of Association*, Thousand Oaks, Sage (Quantitative Applications in the Social Sciences).
- LIJPHART A., 1971, « Comparative politics and the comparative method », *American Political Science Review*, 65, 03, p. 682-693.

- LIPPE T. VAN DER, FODOR É., 1998, « Changes in Gender Inequality in Six Eastern European Countries », *Acta Sociologica*, 41, 2-3, p. 131-149.
- LIPSET S.M., ZETTERBERG H.L., 1959, « Social Mobility in Industrial Society », dans LIPSET S.M., BENDIX R. (dirs.), *Social Mobility in Industrial Society*, Berkeley, University of California Press, p. 11-75.
- LÓPEZ-RUIZ L., ESTEVE A., CABRÉ A., 2009, « Consensual Unions and Marriages in America Latina: two patterns of educative homogamy? », *Papeles de Población*, 60, p. 5-34.
- LORDON F., 2011, *L'intérêt souverain. Essai d'anthropologie économique spinoziste*, Paris, La Découverte.
- LÖWY I., 2006, *L'emprise du genre. Masculinité, féminité, inégalité*, Paris, La Dispute.
- LUMLEY T., 2014, *survey: analysis of complex survey samples*, R package, version 3.30-3.
- LUMLEY T., SCOTT A., 2015, « AIC and BIC for Modeling with Complex Survey Data », *Journal of Survey Statistics and Methodology*, 3, 1, p. 1-18.
- LUNN M., MCNEIL D., 1995, « Applying Cox Regression to Competing Risks », *Biometrics*, 51, 2, p. 524-532.
- MACFARLANE A., 1986, *Marriage and love in England: Modes of reproduction 1300-1840*, New York, Blackwell.
- MACH B.W., KRYSZCZUK M.D., 2008, « The ISCED-97 in the Polish context », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 226-235.
- MÄENPÄÄ E., 2014, « Homogamy in Educational Level and Parental Social Class in Finland: A Log-linear Analysis », *European Sociological Review*, 31, 3, p. 253-267.
- MÄENPÄÄ E., JALOVAARA M., 2014a, « Homogamy in socio-economic background and education, and the dissolution of cohabiting unions », *Demographic Research*, 30, p. 1769-1792.
- MÄENPÄÄ E., JALOVAARA M., 2014b, « Achievement Replacing Ascription? Changes in Homogamy in Education and Social Class Origins », *Stockholm Research Reports in Demography*, 23, Stockholm, Stockholm University.
- MAILLOCHON F., 2003, « Le jeu de l'amour et de l'amitié au lycée: mélange des genres », *Travail, genre et sociétés*, 9, p. 111-135.
- MANDEL H., 2009, « Configurations of gender inequality: the consequences of ideology and public policy », *The British Journal of Sociology*, 60, 4, p. 693-719.
- MARCHAND O., THÉLOT C., 1997, *Le travail en France: 1800-2000*, Paris, Nathan (Essais & Recherches).
- MARE R.D., 1991, « Five Decades of Educational Assortative Mating », *American Sociological Review*, 56, 1, p. 15-32.
- MARE R.D., SCHWARTZ C.R., 2006, « Educational Assortative Mating and the Family Background of the Next Generation », *Sociological Theory and Methods*, 21, 2, p. 253-278.

- MARRY C., KIEFFER A., BRAUNS H., STEINMANN S., 1998, « France–Allemagne: Inégales avancées des femmes. Évolutions comparées de l'éducation et de l'activité des femmes de 1971 à 1991 », Working Paper, 26, Mannheim, MZER.
- MARUANI M., 2000, *Travail et emploi des femmes*, Paris, La Découverte (Repères).
- MARUANI M., 2002, *Les Mécomptes du chômage*, Paris, Bayard.
- MASSE B., 2002, « Rites scolaires et rites festifs: les “manières de boire” dans les grandes écoles », *Sociétés contemporaines*, 47, 3, p. 101-129.
- MASSON L., 2013, « Avez-vous eu des enfants? Si oui, combien? », *France, portrait social 2013*, Paris, Insee (Insee Références), p. 92-109.
- MATOSSIAN M., 1976, « Review: Edward Shorter. The Making of the Modern Family, and Charles E. Rosenberg, editor. The Family in History. », *The American Historical Review*, 81, 4, p. 822-823.
- MAUGER G., 2006, « Les transformations des classes populaires en France depuis 30 ans », dans LOJKINE J., COURS-SALIES P., VAKALOULIS M. (dirs.), *Nouvelles luttes de classes*, Paris, Presses universitaires de France (Actuel Marx Confrontation), p. 29-42.
- MAUGER G., 2013, « Bourdieu et les classes populaires. L'ambivalence des cultures dominées », dans COULANGEON P., DUVAL J. (dirs.), *Trente ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte (Recherches), p. 243-254.
- MAURIN É., 2004, *Le ghetto français. Enquête sur le séparatisme social*, Paris, Seuil (La République des idées).
- MENDRAS H., 1988, *La Seconde Révolution française, 1965-1984*, NRF, Paris, Gallimard.
- MERLE P., 2000, « Le concept de démocratisation de l'institution scolaire: Une typologie et sa mise à l'épreuve », *Population*, 55, 1, p. 15-50.
- MERLLIÉ D., COUSQUER J.-Y., 1980, « Mariage et relations familiales dans l'aristocratie rurale: deux entretiens », *Actes de la recherche en sciences sociales*, 31, 1, p. 22-34.
- MERLLIÉ D., PRÉVOT J., 1997, *La mobilité sociale*, Paris, La Découverte (Repères).
- MERON M., 2008, « Femmes et hommes dans l'emploi: permanences et évolutions », *L'emploi, nouveaux enjeux*, Édition 2008, Paris, Insee (Insee Références), p. 85-93.
- MERON, M., AMAR, M. (dirs.), 2014, *Final Report of the European Statistical System Network on the Harmonisation and Implementation of a European Socio-Economic Classification: European Socio-Economic Groups (ESeG)*, Kirchberg, Eurostat.
- MERON M., MARUANI M., 2012, *Un siècle de travail des femmes en France. 1901-2011*, Paris, La Découverte.
- MERON M., OKBA M., VINEY X., 2006, « Les femmes et les métiers: vingt ans d'évolutions contrastées », *Données sociales 2006*, Paris, Insee, p. 225-234.
- MEVIK B.-H., WEHRENS R., LILAND K.H., 2013, *pls: Partial Least Squares and Principal Component regression*, R package, version 2.4-3.
- MEYER D., ZEILEIS A., HORNIK K., 2014, *vcd: Visualizing Categorical Data*, R package, version 1.3-2.

- MICHELAT G., SIMON M., 2004, *Les Ouvriers et la politique: Permanence, ruptures, réalignements*, Paris, Presses de Sciences Po.
- MIGNOT J.-F., 2010, « L'écart d'âge entre conjoints », *Revue française de sociologie*, 51, 2, p. 281-320.
- MOCANU C., 2008, « Romania », dans KOGAN I., GEBEL M., NOELKE C. (dirs.), *Europe Enlarged: A Handbook of Education, Labour and Welfare Regimes in Central and Eastern Europe*, Bristol, Policy Press, p. 295-322.
- MOLINIER J., 1977, « L'évolution de la population agricole du XVIII^e siècle à nos jours », *Économie et statistique*, 91, p. 79-84.
- MONAGHAN D., 2014, « The Impact of Economic Inequality and Gender Parity on Educational Assortative Mating: Evidence from the Luxembourg Income Study », LIS Working Paper, 607, Luxembourg, Luxembourg Income Study.
- MONAGHAN D., 2015, « Income inequality and educational assortative mating: Evidence from the Luxembourg Income Study », *Social Science Research*, 52, p. 253-269.
- MORIN T., 2014, « Écarts de revenus au sein des couples: Trois femmes sur quatre gagnent moins que leur conjoint », *Insee Première*, 1492.
- MOUNT F., 1982, *La famille subversive*, Bruxelles, Mardaga.
- MUHSAM H.V., 1974, « The marriage squeeze », *Demography*, 11, 2, p. 291-299.
- MÜLLER W., KOGAN I., 2010, « Education », dans IMMERSALL S., THERBORN G. (dirs.), *Handbook of European Societies: Social Transformation in the 21st Century*, Berlin-New York, Springer, p. 217-290.
- MURTAUGH P.A., 2014, « In defense of P values », *Ecology*, 95, 3, p. 611-617.
- NAUZE-FICHET E., GONZALEZ-DEMICHÉL C., 2003, « Les contours de la population active: aux frontières de l'emploi, du chômage et de l'inactivité », *Économie et statistique*, 362, p. 85-103.
- NÍ BHROLCHÁIN M., SIGLE-RUSHTON W., 2005, « L'offre de conjoints potentiels en Grande-Bretagne et aux États-Unis », *Population*, 60, 1, p. 39-70.
- NICOLE-DRANCOURT C., 1989, « Stratégies professionnelles et organisation des familles », *Revue française de sociologie*, 30, 1, p. 57-80.
- OCDE, 2008, *Croissance et inégalités: Distribution des revenus et pauvreté dans les pays de l'OCDE*, Paris, Organisation de coopération et de développement économiques.
- OCDE, 2011, *Toujours plus d'inégalité: Pourquoi les écarts de revenus se creusent*, Paris, Organisation de coopération et de développement économiques.
- OPPENHEIMER V.K., 1988, « A Theory of Marriage Timing », *The American Journal of Sociology*, 94, 3, p. 563-591.
- OPPENHEIMER V.K., 1994, « Women's Rising Employment and the Future of the Family in Industrial Societies », *Population and Development Review*, 20, 2, p. 293-342.
- OPPENHEIMER V.K., 1995, « The Role of Women's Economic Independence in Marriage Formation: A Skeptic's Response to Annemette Sørensen's Remarks », dans BLOSSFELD

- H.-P. (dir.), *The New Role of Women: Family Formation in Modern Societies*, Boulder, Westview Press, p. 236-243.
- ORTIZ L., 2008, « Evaluation of the ISCED-97 for the Spanish system of education », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 125-148.
- PAILHÉ A., MORTELMANS D., CASTRO T., CORTINA TRILLA C., DIGOIX M., FESTY P., KRAPF S., KREYENFELD M., LYSSENS-DANNEBOOM V., MARTÍN-GARCÍA T., RAULT W., THÉVENON O., TOULEMON L., 2014, « State-of-the-art report: Changes in the life course », Working Paper Series, 6, Families And Societies European Union's Seventh Framework Project.
- PALHETA U., 2011, « Le collège divise. Appartenance de classe, trajectoires scolaires et enseignement professionnel », *Sociologie*, 2, 4, p. 363-386.
- PAN KÉ SHON J.-L., 1998, « D'où sont mes amis venus ? », *Insee Première*, 613.
- PARSONS T., BALES R.F., 1955, *Family, Socialization and Interaction Process*, Glencoe, Free Press.
- PAUGAM S., 1991, *La disqualification sociale. Essai sur la nouvelle pauvreté*, Paris, Presses Universitaires de France (Sociologies).
- PAUGAM S., 2000, *Le Salarié de la précarité*, Paris, Presses Universitaires de France (Le lien social).
- PENCAVEL J., 1998, « Assortative Mating by Schooling and the Work Behavior of Wives and Husbands », *The American Economic Review*, 88, 2, p. 326-329.
- PENCAVEL J., 2006, « A life cycle perspective on changes in earnings inequality among married men and women », *The Review of Economics and Statistics*, 88, 2, p. 232-242.
- PEUGNY C., 2008, « Éducation et mobilité sociale: la situation paradoxale des générations nées dans les années 1960 », *Économie et statistique*, 410, p. 23-45.
- PEUGNY C., 2011, « Les enfants de cadre: fréquence et ressorts du déclassement », dans BOUFFARTIGUE P., GADEA C., POCHIC S. (dirs.), *Cadres, classes moyennes: vers l'éclatement ?*, Recherches, Paris, Armand Colin, p. 173-186.
- PIERRU E., SPIRE A., 2008, « Le crépuscule des catégories socioprofessionnelles », *Revue française de science politique*, 58, 3, p. 457-481.
- PIKETTY T., 2001, *Les hauts revenus en France au XX^e siècle*, Paris, Grasset.
- PIKETTY T., 2003, « Income Inequality in France, 1901-1998 », *Journal of Political Economy*, 111, 5, p. 1004-1042.
- PINÇON M., PINÇON-CHARLOT M., 1989, *Dans les beaux quartiers*, Paris, Seuil.
- PLACE D., VINCENT B., 2009, « L'influence des caractéristiques sociodémographiques sur les diplômes et les compétences », *Économie et statistique*, 424, p. 125-147.
- PONTHIEUX S., SCHREIBER A., 2006, « Dans les couples de salariés, la répartition du travail domestique reste inégale », *Données sociales 2006*, Paris, Insee, p. 43-51.
- PÖTTER U., BLOSSFELD H.-P., 2001, « Causal Inference from Series of Events », *European Sociological Review*, 17, 1, p. 21-32.

- POULLAOUËC T., 2010, *Le diplôme, arme des faibles. Les familles ouvrières et l'école*, Paris, La Dispute (L'enjeu scolaire).
- PRÉTECEILLE E., 2006, « La ségrégation sociale a-t-elle augmenté ? », *Sociétés contemporaines*, 62, 2, p. 69-93.
- PRIOUX F., 2003, « L'âge à la première union en France: une évolution en deux temps », *Population*, 58, 4, p. 623-644.
- PRIOUX F., 2005, « Mariage, vie en couple et rupture d'union », 122, 2, p. 38-50.
- PRIOUX F., MAZUY M., BARBIERI M., 2011, « L'évolution démographique récente en France: les adultes vivent moins souvent en couple », *Population*, 65, 3, p. 421-474.
- PUJOL J., TOMASINI M., 2009, « Les inégalités de niveaux de vie entre 1996 et 2007 », *Insee Première*, 1266.
- PUTTER H., FIOCCO M., GESKUS R.B., 2007, « Tutorial in biostatistics: competing risks and multi-state models », *Statistics in Medicine*, 26, 11, p. 2389-2430.
- PUUR A., RAHNU L., MASLAUSKAITE A., STANKUNIENE V., ZAKHAROV S., 2012, « Transformation of Partnership Formation in Eastern Europe: The Legacy of the Past Demographic Divide », *Journal of Comparative Family Studies*, 43, 3, p. 389-417.
- QIAN Z., 1998, « Changes in Assortative Mating: The Impact of Age and Education, 1970-1990 », *Demography*, 35, 3, p. 279-292.
- QIAN Z., PRESTON S.H., 1993, « Changes in American Marriage, 1972 to 1987: Availability and Forces of Attraction by Age and Education », *American Sociological Review*, 58, 4, p. 482-495.
- RAFTERY A.E., 1986, « Choosing models for cross-classifications », *American Sociological Review*, 51, 1, p. 145-146.
- RAFTERY A.E., 1995, « Bayesian Model Selection in Social Research », *Sociological Methodology*, 25, p. 111-196.
- RAYMO J.M., XIE Y., 2000, « Temporal and Regional Variation in the Strength of Educational Homogamy », *American Sociological Review*, 65, 5, p. 773-781.
- R CORE TEAM, 2014, *R: A Language and Environment for Statistical Computing*, Vienna, version 3.1.2.
- REED D., CANCIAN M., 2001, « Sources of inequality: measuring the contributions of income sources to rising family income inequality », *Review of Income and Wealth*, 47, 3, p. 321-333.
- REED D., CANCIAN M., 2012, « Rising Family Income Inequality: The Importance of Sorting », *Journal of Income Distribution*, 21, 2, p. 3-14.
- RÉGNIER-LOILIER A., BEAUJOUAN É., VILLENEUVE-GOKALP C., 2009, « Neither single, nor in a couple: A study of living apart together in France », *Demographic research*, 21, p. 75-108.
- REHER D.S., 1998, « Family Ties in Western Europe: Persistent Contrasts », *Population and Development Review*, 24, 2, p. 203-234.

- RICROCH L., 2012, « En 25 ans, moins de tâches domestiques pour les femmes, l'écart de situation avec les hommes se réduit », *Regards sur la parité 2012*, Paris, Insee (Insee Références), p. 67-80.
- ROBERT-BOBÉE I., MAZUY M., 2005, « Calendriers de constitution des familles et âge de fin des études », dans LEFÈVRE C., FILHON A. (dirs.), *Histoires de familles, histoires familiales : les résultats de l'enquête famille de 1999*, Paris, Ined (Cahiers), p. 175-200.
- RÓBERT P., 2010, « Stratification and Social Mobility », dans IMMERFALL S., THERBORN G. (dirs.), *Handbook of European Societies: Social Transformation in the 21st Century*, Berlin-New York, Springer, p. 499-536.
- RODRÍGUEZ S., 2014, « Pautas y tendencias de homogamia educacional relativa en Argentina a comienzos del siglo XXI », *Entramados y Perspectivas*, 2, p. 99-126.
- ROSE D., HARRISON E., 2009, *Social Class in Europe*, London-New York, Routledge.
- ROSE E., 2004, « Education and Hypergamy in Marriage Markets », CSDE Working Paper, 04-03, Seattle, Center for Studies in Demography and Ecology, University of Washington.
- ROUSSEL L., 1971, « La nuptialité en France. Précocité et intensité suivant les régions et les catégories socio-professionnelles », *Population*, 26, 6, p. 1029-1055.
- ROUSSEL, L. (dir.), 1975, *Le Mariage dans la société française contemporaine : faits de population, données d'opinion*, Paris, Presses Universitaires de France (Travaux et documents de l'Ined).
- ROUSSEL L., 1978, « La cohabitation juvénile en France », *Population*, 33, 1, p. 15-42.
- RÓŻAŃSKA-PUTEK J., JAPPENS M., WILLAERT D., VAN BAVEL J., 2009, « Recoding the Regions of the European Social Survey into the NUTS 1 Regional Classification. Illustration: regional indicators of intergenerational solidarity », Interface Demography Working Paper, Brussels, Vrije Universiteit Brussel.
- RUDAS T., 1998, *Odds Ratios in the Analysis of Contingency Tables*, Thousand Oaks, Sage (Quantitative Applications in the Social Sciences).
- SANTELLI E., COLLET B., 2011, « De l'endogamie à l'homogamie socio-ethnique: Réinterprétations normatives et réalités conjugales des descendants d'immigrés maghrébins, turcs et africains sahéliens », *Sociologie et sociétés*, 43, 2, p. 329-354.
- SCHNAPPER D., 1989, « Rapport à l'emploi, protection sociale et statuts sociaux », *Revue française de sociologie*, 30, 1, p. 3-29.
- SCHNEIDER, S.L. (dir.), 2008a, *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES.
- SCHNEIDER S.L., 2008b, « The application of the ISCED-97 to the UK's educational qualifications », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 281-300.
- SCHNEIDER S.L., 2008c, « Applying the ISCED-97 to the German educational qualifications », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An*

- Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 76-102.
- SCHNEIDER S.L., 2009, *Confusing credentials: the cross-nationally comparable measurement of educational attainment*, Thèse de doctorat, Oxford, University of Oxford.
- SCHNEIDER S.L., KOGAN I., 2008, « The International Standard Classification of Education 1997: challenges in the application to national data and the implementation in cross-national surveys », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 13-46.
- SCHWARTZ C.R., 2010, « Earnings Inequality and the Changing Association between Spouses' Earnings », *American Journal of Sociology*, 115, 5, p. 1524-1557.
- SCHWARTZ C.R., 2013, « Trends and Variation in Assortative Mating: Causes and Consequences », *Annual Review of Sociology*, 39, 1, p. 451-470.
- SCHWARTZ C.R., MARE R.D., 2005, « Trends in Educational Assortative Marriage from 1940 to 2003 », *Demography*, 42, 4, p. 621-646.
- SCHWARZ G., 1978, « Estimating the dimension of a model », *The Annals of Statistics*, 6, 2, p. 461-464.
- SCHWEITZER S., 2002, *Les femmes ont toujours travaillé: Une histoire du travail des femmes aux XIX^e et XX^e siècles*, Paris, Odile Jacob.
- SECCOMBE W., 1993, *Weathering the Storm: Working-Class Families from the Industrial Revolution to the Fertility Decline*, London, Verso.
- SEGALEN M., [1981] 2010, *Sociologie de la famille*, 7^e édition, Paris, Armand Colin.
- SEGALEN M., JACQUARD A., 1971, « Choix du conjoint et homogamie », *Population*, 26, 3, p. 487-498.
- SELVIN H.C., 1957, « A Critique of Tests of Significance in Survey Research », *American Sociological Review*, 22, 5, p. 519-527.
- SHALEV M., 2007, « Limits and alternatives to multiple regression in comparative research », *Comparative Social Research*, 24, 3, p. 261-308.
- SHORROCKS A.F., 1980, « The class of additively decomposable inequality measures », *Econometrica*, 48, 3, p. 613-625.
- SHORTER E., [1975] 1977, *Naissance de la famille moderne: XVIII^e-XX^e siècle*, Seuil (L'univers historique).
- SIMMEL G., 1897, « Comment les formes sociales se maintiennent », *L'Année sociologique*, 1, p. 71-109.
- SIMMEL G., [1907] 1998, *Les pauvres*, Paris, Presses Universitaires de France (Quadrige).
- SIMMEL G., [1908a] 1999, « Le croisement des cercles sociaux », *Sociologie. Études sur les formes de la socialisation*, Paris, Presses Universitaires de France (Sociologies), p. 407-252.

- SIMMEL G., [1908b] 1999, « L'élargissement du groupe et le développement de l'individualité », *Sociologie. Études sur les formes de la socialisation*, Paris, Presses Universitaires de France (Sociologies), p. 685-746.
- SINGLY F. DE, 1982, « Mariage, dot scolaire et position sociale », *Économie et statistique*, 142, p. 7-20.
- SINGLY F. DE, 1984, « Les manœuvres de séduction: une analyse des annonces matrimoniales », *Revue française de sociologie*, 25, 4, p. 523-559.
- SINGLY F. DE, 1987a, « Théorie critique de l'homogamie », *L'Année sociologique*, 37, p. 181-205.
- SINGLY F. DE, 1987b, *Fortune et infortune de la femme mariée*, Paris, Presses Universitaires de France.
- SINGLY F. DE, 1988, « L'amour, un bien privé, un mal public? », *Revue française des affaires sociales*, 42, 2, p. 129-142.
- SINGLY F. DE, [1993] 2014, *Sociologie de la famille contemporaine*, 5^e édition, Paris, Armand Colin (128).
- SINGLY F. DE, 1996a, « Trois thèses sur la famille contemporaine », dans LE GALL D., MARTIN C. (dirs.), *Familles et politiques sociales. Dix questions sur le lien familial contemporain*, Paris, L'Harmattan (Logiques sociales), p. 57-69.
- SINGLY F. DE, 1996b, *Le Soi, le couple et la famille*, Paris, Nathan.
- SINGLY F. DE, GLAUDE M., 1986, « L'organisation domestique: pouvoir et négociation », *Économie et statistique*, 187, p. 3-30.
- SMEENK W.H., 1998, *Opportunity and marriage: The impact of individual resources and marriage market structure on first marriage timing and partner choice in the Netherlands*, Amsterdam, Thela Thesis.
- SMITS J., 2003, « Social Closure Among the Higher Educated: Trends in Educational Homogamy in 55 Countries », *Social Science Research*, 32, 2, p. 251-277.
- SMITS J., PARK H., 2009, « Five Decades of Educational Assortative Mating in 10 East Asian Societies », *Social Forces*, 88, 1, p. 227-255.
- SMITS J., ULTEE W., LAMMERS J., 1998, « Educational Homogamy in 65 Countries: An Explanation of Differences in Openness Using Country-Level Explanatory Variables », *American Sociological Review*, 63, 2, p. 264-285.
- SMITS J., ULTEE W., LAMMERS J., 1999, « Occupational Homogamy in Eight Countries of the European Union, 1975-89 », *Acta Sociologica*, 42, 1, p. 55-68.
- SMITS J., ULTEE W., LAMMERS J., 2000, « More or Less Educational Homogamy? A Test of Different Versions of Modernization Theory Using Cross-Temporal Evidence for 60 Countries », *American Sociological Review*, 65, 5, p. 781-788.
- SOBOTKA T., TOULEMON L., 2008, « Changing family and partnership behaviour: Common trends and persistent diversity across Europe », *Demographic Research*, 19, p. 85-138.

- SOLARD J., 2010, « Les très hauts revenus: des différences de plus en plus marquées entre 2004 et 2007 », *Les revenus et le patrimoine des ménages 2010*, Paris, Insee (Insee Références), p. 45-64.
- SOLAZ A., 2000, « Une réflexion économique sur le lien famille-chômage et la constitution du couple », *Recherches et Prévisions*, 60, p. 19-34.
- SOLAZ A., 2013, « Chômage et vie en couple: quelles relations? », *Regards croisés sur l'économie*, 13, 1, p. 67-80.
- SOLT F., à paraître, « The Standardized World Income Inequality Database », *Social Science Quarterly*.
- SOMBART W., [1906] 1992, *Pourquoi le socialisme n'existe-t-il pas aux États-Unis?*, Paris, Presses Universitaires de France.
- SOROKIN P., [1927] 1998, *Social Mobility*, London, Routledge.
- SOUTH S.J., 1991, « Sociodemographic Differentials in Mate Selection Preferences », *Journal of Marriage and Family*, 53, 4, p. 928-940.
- SPRECHER S., SULLIVAN Q., HATFIELD E., 1994, « Mate selection preferences: Gender differences examined in a national sample », *Journal of Personality and Social Psychology*, 66, 6, p. 1074-1080.
- STANDING G., 2011, *The Precariat: The New Dangerous Class*, London, Bloomsbury Academic.
- STONE L., 1977, *The Family, Sex and Marriage in England 1500-1800*, London, Weidenfeld and Nicolson.
- STRAKOVÁ J., 2008, « The Czech educational system and evaluation of the ISCED-97 implementation », dans SCHNEIDER S.L. (dir.), *The International Standard Classification of Education. An Evaluation of Content and Criterion Validity for 15 European Countries*, Mannheim, MZES, p. 216-225.
- TENENHAUS M., 1998, *La régression PLS: théorie et pratique*, Paris, Technip.
- TERRAIL J.-P., 1995, « L'essor contemporain de la scolarisation des filles et son interprétation », dans EPHESIA (dir.), *La place des femmes. Les enjeux de l'identité et de l'égalité au regard des sciences sociales*, Paris, La Découverte (Recherches), p. 586-590.
- TESTENOIRE A., 2008, « Carrières féminines, résistances masculines: couples à hypogamie féminine », dans KERGOAT D., GUICHARD-CLAUDIC Y., VILBROD A. (dirs.), *L'inversion du genre. Quand les métiers masculins se conjuguent au féminin... et réciproquement*, Rennes, Presses Universitaires de Rennes.
- THÉLOT C., 1982, *Tel père, tel fils? Position sociale et origine familiale*, Paris, Dunod.
- THÉLOT C., 1983, « L'évolution de la mobilité sociale dans chaque génération », *Économie et statistique*, 161, p. 3-21.
- THERNEAU T.M., 2014, *survival: A Package for Survival Analysis in S*, R package, version 2.37-7.
- THERNEAU T.M., GRAMBSCH P.M., 2000, *Modeling Survival Data: Extending the Cox Model*, New York, Springer.

- THORNTON A., 2001, « The Developmental Paradigm, Reading History Sideways, and Family Change », *Demography*, 38, 4, p. 449-465.
- TOCQUEVILLE A. DE, [1835] 1961, *De la démocratie en Amérique*, I, Paris, Gallimard (Folio Histoire).
- TOCQUEVILLE A. DE, [1840] 1961, *De la démocratie en Amérique*, II, Paris, Gallimard (Folio Histoire).
- TODD E., 1990, *L'invention de l'Europe*, Paris, Seuil.
- TODD E., 2011, *L'origine des systèmes familiaux*, Tome I. L'Eurasie, Paris, Gallimard (NRF Essais).
- TODD P.M., BILLARI F.C., SIMÃO J., 2005, « Aggregate Age-at-Marriage Patterns from Individual Mate-Search Heuristics », *Demography*, 42, 3, p. 559-574.
- TODD P.M., MILLER G.F., 1999, « From Pride and Prejudice to Persuasion: Satisficing in Mate Search », dans TODD P.M., GIGENZERER G., ABC RESEARCH GROUP (dirs.), *Simple Heuristics That Make Us Smart*, Oxford, Oxford University Press (Evolution and Cognition), p. 287-308.
- TORCHE F., 2010, « Educational assortative mating and economic inequality: A comparative analysis of three Latin American countries », *Demography*, 47, 2, p. 481-502.
- TOULEMON L., 1996, « La cohabitation hors mariage s'installe dans la durée », *Population*, 51, 3, p. 675-715.
- TOULEMON L., 1998, « Situation professionnelle et comportements familiaux des indépendants », *Économie et statistique*, 319, p. 29-51.
- TREIMAN D.J., 1970, « Industrialization and Social Stratification », *Sociological Inquiry*, 40, 2, p. 207-234.
- TREIMAN D.J., GANZEBOOM H.B., 2000, « The Fourth Generation of Comparative Stratification Research », dans QUAH S.R., SALES A. (dirs.), *The International Handbook of Sociology*, London, Sage, p. 122-150.
- TURNER H., FIRTH D., 2014, *Generalized nonlinear models in R: An overview of the gnm package*, R package, version 1.0-7.
- ULTEE W., 2007, « Stratification Systems: Openness », dans RITZER G. (dir.), *Blackwell Encyclopedia of Sociology*, Malden, Blackwell, p. 4819-4820.
- ULTEE W., DESSENS J., JANSEN W., 1988, « Why does unemployment come in couples? An analysis of (un)employment and (non)employment homogamy tables for Canada, the Netherlands and the United States in the 1980s », *European Sociological Review*, 4, 2, p. 111-122.
- ULTEE W., LUIJKX R., 1990, « Educational Heterogamy and Father-to-Son Occupational Mobility in 23 Industrial Nations: General Societal Openness or Compensatory Strategies of Reproduction? », *European Sociological Review*, 6, 2, p. 125-149.
- UUNK W.J.G., 1996, *Who Marries Whom? The Role of Social Origin, Education and High Culture in Mate Selection of Industrial Societies During the Twentieth Century*, Thèse de doctorat, Nijmegen, University of Nijmegen.

- VALLET L.-A., 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*, 27, 4, p. 655-696.
- VALLET L.-A., 1999, « Quarante années de mobilité sociale en France. L'évolution de la fluidité sociale à la lumière de modèles récents », *Revue française de sociologie*, 40, 1, p. 5-64.
- VALLET L.-A., 2001, « Stratification et mobilité sociales: la place des femmes », dans LAUFER J., MARRY C., MARUANI M. (dirs.), *Masculin-Féminin: questions pour les sciences de l'homme*, Paris, Presses Universitaires de France, p. 81-97.
- VALLET L.-A., 2004, « Change in intergenerational class mobility in France from the 1970s to the 1990s and its explanation: An analysis following the CASMIN approach », *Social Mobility in Europe*, Oxford, Oxford University Press, p. 115-147.
- VALLET L.-A., 2007, « Sur l'origine, les bonnes raisons de l'usage, et la fécondité de l'odds ratio », *Courrier des statistiques*, 121-122, p. 59-65.
- VALLET L.-A., 2011, « L'évolution du recrutement social des cadres en France (1953-2003): l'ouverture plutôt que la clôture sociale? », dans BOUFFARTIGUE P., GADEA C., POCHIC S. (dirs.), *Cadres, classes moyennes: vers l'éclatement?*, Paris, Armand Colin (Recherches), p. 166-172.
- VALLET L.-A., SELZ M., 2007, « Évolution historique de l'inégalité des chances devant l'école: des méthodes et des résultats revisités », *Éducation & formations*, 74, p. 65-74.
- VANDERSCHULDEN M., 2006a, « Homogamie socioprofessionnelle et ressemblance en termes de niveau d'études: constat et évolution au fil des cohortes d'unions », *Économie et statistique*, 398-399, p. 33-58.
- VANDERSCHULDEN M., 2006b, « L'écart d'âge entre conjoints s'est réduit », *Insee Première*, 1073.
- VANDERSCHULDEN M., 2006c, « Position sociale et choix du conjoint: des différences marquées entre hommes et femmes », *Données sociales 2006*, Paris, Insee, p. 33-42.
- VERBAKEL E., 2008, *The Partner as Resource or Restriction? Labour market careers of husbands and wives and the consequences for inequality between couples*, Thèse de doctorat, Nijmegen, Radboud University.
- VERBAKEL E., LUIJKX R., GRAAF P.M. DE, 2008, « The association between husbands' and wives' labor market positions in the Netherlands », *Research in Social Stratification and Mobility*, 26, 3, p. 257-276.
- VERDUGO G., FRAISSE H., HORNY G., 2012, « Évolution des inégalités salariales en France: le rôle des effets de composition », *Revue économique*, 63, 6, p. 1081-1112.
- VEYNE P., 1978, « La famille et l'amour sous le Haut-Empire romain », *Annales. Histoire, Sciences Sociales*, 33, 1, p. 35-63.
- VILLAC M., 1984, « Structures familiales et milieux sociaux », *Économie et statistique*, 171, p. 135-151.
- VILLENEUVE-GOKALP C., 1990, « Du mariage aux unions sans papiers: histoire récente des transformations conjugales », *Population*, 45, 2, p. 265-297.

- VILLUND O., 2010, « Household data in the Norwegian Labour Force Survey », *Documents*, 46, Oslo, Statistics Norway.
- VULTUR M., 2010, « La précarité : un “concept fantôme” dans la réalité mouvante du monde du travail », *SociologieS*, disponible sur <https://sociologies.revues.org/3287>, mis en ligne le 27 septembre 2010.
- WAGNER A.-C., 2008, « Mariages assortis et logiques de l'entre-soi dans l'aristocratie et dans la haute bourgeoisie », *Migrations Société*, 119, p. 229-242.
- WEAKLIEM D.L., 1999, « A Critique of the Bayesian Information Criterion for Model Selection », *Sociological Methods & Research*, 27, 3, p. 359-397.
- WEBER M., [1922] 1968, *Economy and Society. An Outline of Interpretive Sociology*, Berkeley, University of California Press.
- WEBER M., [1922] 1971, *Économie et société*, Paris, Plon.
- WESTERN B., BLOOME D., PERCHESKI C., 2008, « Inequality among American families with children, 1975 to 2005 », *American Sociological Review*, 73, 6, p. 903-920.
- WIDMER É., 1993, *De cœur et de raison. Le choix du conjoint à Genève au XIX^e siècle*, Cahiers, Société d'Histoire et d'Archéologie de Genève.
- WIDMER É.D., 2014, « Un demi-siècle de pluralisation des formes familiales, et après ? », dans SERVAIS P. (dir.), *Regards sur la famille, le couple et la sexualité : un demi-siècle de mutations*, Louvain-La-Neuve, Academia, p. 35-46.
- WILKINSON G.N., ROGERS C.E., 1973, « Symbolic Description of Factorial Models for Analysis of Variance », *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 22, 3, p. 392-399.
- WINKLER-DWORAK M., TOULEMON L., 2007, « Gender Differences in the Transition to Adulthood in France: Is There Convergence Over the Recent Period? », *European Journal of Population*, 23, 3-4, p. 273-314.
- WONG R.S.-K., 2001, « Multidimensional Association Models », *Sociological Methods & Research*, 30, 2, p. 197-240.
- WONG R.S.-K., 2003, « To see or not to see: Another look at research on temporal trends and cross-national differences in educational homogamy », *Taiwanese Journal of Sociology*, 31, p. 47-91.
- WONG R.S.-K., 2010, *Association models*, Thousand Oaks, Sage (Quantitative Applications in the Social Sciences).
- WORNER S.M., 2006, « The Effects of Assortative Mating on Income Inequality: A Decompositional Analysis », CEPR Discussion Paper, 538, Canberra, Centre for Economic Policy Research, Research School of Economics, Australian National University.
- XIE Y., 1992, « The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables », *American Sociological Review*, 57, 3, p. 380-395.
- XIE Y., 1999, « The Tension Between Generality and Accuracy », *Sociological Methods & Research*, 27, 3, p. 428-435.

- YAISH M., KATZ-GERRO T., 2010, « Disentangling 'Cultural Capital': The Consequences of Cultural and Economic Resources for Taste and Participation », *European Sociological Review*, 28, 2, p. 169-185.
- YAMAGUCHI K., 1990, « Some Models for the Analysis of Asymmetric Association in Square Contingency Tables with Ordered Categories », *Sociological Methodology*, 20, p. 181-212.
- YAMAGUCHI K., 1991, *Event History Analysis*, Newbury Park, CA, Sage (Applied Social Research Methods).
- YAMAGUCHI K., 1992, « Accelerated Failure-Time Regression Models with a Regression Model of Surviving Fraction: An Application to the Analysis of "Permanent Employment" in Japan », *Journal of the American Statistical Association*, 87, 418, p. 284-292.
- ZARCA B., 1990, « La division du travail domestique: poids du passé et tensions au sein du couple », *Économie et statistique*, 228, p. 29-40.

LISTE DES FIGURES

Figure 1.1 : Évolution de la proportion d'hommes et de femmes en couple.....	79
Figure 1.2 : Évolution de la structure par diplôme des 30-59 ans (proportions cumulées).....	94
Figure 1.3 : Évolution du taux d'endogamie de diplôme des 30-59 ans attendu en situation de mise en couple aléatoire.....	95
Figure 1.4 : Taux d'endogamie, d'hypergamie et d'hypogamie de diplôme observés et attendus en situation de choix aléatoire du conjoint.....	96
Figure 1.5 : Répartition des couples selon les diplômes des conjoints en 1969-71 et en 2009-11 (% en ligne).....	99
Figure 1.6 : Évolution de la structure par classe sociale des actifs âgés de 30 à 59 ans.....	105
Figure 1.7 : Évolution du taux d'endogamie de classe sociale des actifs de 30 à 59 ans attendu en situation de mise en couple aléatoire.....	107
Figure 1.8 : Taux d'endogamie de classe sociale observés et attendus en situation de choix aléatoire du conjoint.....	108
Figure 1.9 : Répartition des couples selon la classe EGP des conjoints en 1969-71 et en 2009-11 (% en ligne).....	112
Figure 1.10 : Répartition des couples selon la catégorie socioprofessionnelle des conjoints en 1982-85 et en 2009-11 (% en ligne).....	113
Figure 1.11 : Évolution de la structure par classe sociale d'origine des 30-59 ans.....	115
Figure 1.12 : Taux d'endogamie d'origine sociale observés et attendus en situation de choix aléatoire du conjoint.....	117
Figure 1.13 : Répartition des couples selon la classe EGP d'origine des conjoints en 1982-85 et en 2009-11 (% en ligne).....	119
Figure 2.1 : Évolution de l'intensité de l'homogamie d'après le modèle UNIDIFF.....	148
Figure 2.2 : Odds ratios généralisés mesurant l'endogamie dans les modèles à forme de régression.....	151
Figure 2.3 : Dimensions du modèle RC(3)-L pour l'homogamie de diplôme.....	160
Figure 2.4 : Évolution de l'intensité des dimensions des modèles RC-L.....	163
Figure 2.5 : L'homogamie de diplôme en 1969-1971 et en 2009-2011 selon le modèle RC(2)	164
Figure 2.6 : Les dimensions du modèle RC(3)-L pour la classe sociale des actifs.....	167
Figure 2.7 : L'homogamie de classe sociale des actifs en 1969-1971 et en 2009-2011 selon le modèle RC(2).....	171
Figure 2.8 : L'homogamie de PCS des actifs en 2009-2011 selon le modèle RC(2).....	173
Figure 2.9 : Les dimensions du modèle RC(2)-L pour la classe sociale d'origine.....	176
Figure 2.10 : L'homogamie de classe sociale d'origine en 1982-1985 et en 2009-2011 selon le modèle RC(2).....	177
Figure 3.1 : La croissance des salaires des femmes depuis 1982 à travers trois indicateurs. .	195
Figure 3.2 : Évolution des inégalités de salaire et de la corrélation entre salaires des conjoints depuis 1990.....	198
Figure 3.3 : Évolution du taux d'emploi des hommes et des femmes selon le salaire de leur conjoint.....	202
Figure 3.4 : Évolution de deux rapports interdéciles entre salaires des couples depuis 1990	204
Figure 3.5 : Évolution du coefficient de variation et des rapports D5/D2 et D9/D5 entre 1990	

et 2011 dans les différentes situations contrefactuelles.....	210
Figure 3.6 : Évolution de l'association symétrique et de l'hypergamie parmi les couples bi-actifs entre 1990 et 2011.....	224
Figure 3.7 : Évolution de la proportion de couples formés de deux chômeurs ou précaires entre 1975 et 2011.....	230
Figure 4.1 : Évolution des taux d'endogamie, d'hypergamie et d'hypogamie de classe sociale observés et attendus en situation de mise en couple aléatoire.....	256
Figure 4.2 : Évolution des taux d'endogamie, d'hypergamie et d'hypogamie de classe sociale d'origine observés et attendus en situation de mise en couple aléatoire.....	258
Figure 4.3 : Évolution de l'hypergamie relative de diplôme et de classe sociale entre 1969 et 2011.....	265
Figure 4.4 : Proportion d'hommes et de femmes célibataires selon le diplôme, la classe et l'origine sociales en 1975 ou 1982, en 1990 et en 2011.....	271
Figure 4.5 : La paire de dimensions anti-symétrique pour le diplôme.....	281
Figure 4.6 : Évolution de l'intensité des paires de dimensions anti-symétriques.....	282
Figure 4.7 : La paire de dimensions anti-symétrique pour la classe sociale.....	284
Figure 4.8 : La paire de dimensions anti-symétrique pour la classe sociale d'origine.....	287
Figure 5.1 : Relation entre les diplômes des conjoints au sein des premiers couples selon l'année de naissance (proportions observées et proportions attendues sous l'hypothèse de choix aléatoire du conjoint).....	299
Figure 5.2 : Hypergamie et hypogamie relatives de diplôme au sein des premiers couples selon l'année de naissance dans un modèle log-linéaire.....	303
Figure 5.3 : Distribution des âges de mise en couple (courbe de séjour) pour trois groupes de diplômes au sein de deux générations.....	311
Figure 5.4 : Proportion d'individus qui auront vécu au moins une fois en couple selon l'année de naissance et le diplôme.....	312
Figure 5.5 : Âge médian de première mise en couple selon l'année de naissance et le diplôme.....	316
Figure 5.6 : Chances instantanées de mise en couple pour trois groupes de diplômes au sein de deux générations.....	319
Figure 6.1 : Méthode de comparaison de l'âge de fin de scolarité et de l'âge de mise en couple lorsque les deux événements ont lieu la même année.....	341
Figure 6.2 : Proportion des individus s'étant mis en couple pendant leurs études selon le diplôme et l'année de naissance.....	344
Figure 6.3 : Âge médian de fin d'études selon le diplôme et l'année de naissance.....	347
Figure 6.4 : Temps médian écoulé entre la fin des études et la mise en couple par niveau de diplôme et année de naissance.....	348
Figure 6.5 : Relation entre les diplômes des conjoints selon la scolarisation à la mise en couple.....	353
Figure 6.6 : Évolution avec l'âge des chances de rencontrer un conjoint d'un niveau de diplôme donné.....	370
Figure 6.7 : Évolution avec l'âge de la structure par diplôme finalement obtenu de la population des célibataires (proportions cumulées).....	371
Figure 6.8 : Effet d'une année de plus après la fin des études sur les chances de rencontrer un conjoint d'un niveau de diplôme donné.....	374
Figure 6.9 : Effet de la scolarisation sur les chances de rencontrer un conjoint d'un niveau de diplôme donné par rapport à celles de rencontrer un non diplômé.....	377
Figure 6.10 : Effet de la scolarisation sur les chances de rencontrer un conjoint de même	

diplôme que soi.....	379
Figure 6.11 : Effet d'une année de scolarisation de plus sur les chances de rencontrer un conjoint d'un niveau de diplôme donné.....	381
Figure 6.12 : Effet d'une année de plus depuis la fin des études sur les chances de former un couple endogame.....	383
Figure 7.1 : Taux d'endogamie d'éducation et socioéconomique dans les régions d'Europe	424
Figure 7.2 : Taux d'endogamie socioéconomique et proportion de couples de deux indépendants dans les régions d'Europe.....	425
Figure 7.3 : Distribution géographique de l'homogamie relative d'éducation et de groupe socioéconomique en Europe.....	429
Figure 7.4 : Dimensions du modèle à forme de régression multiple pour l'homogamie d'éducation.....	432
Figure 7.5 : Structure de l'homogamie d'éducation dans quatre régions-types.....	436
Figure 7.6 : Dimensions du modèle à forme de régression multiple pour l'homogamie socioéconomique.....	439
Figure 7.7 : Structure de l'homogamie socioéconomique dans quatre régions-types.....	440
Figure 7.8 : Rapport d'hypergamie d'éducation et socioéconomique dans les régions d'Europe.....	444
Figure 7.9 : Géographie de l'intensité de l'hypergamie relative d'éducation et socioéconomique en Europe.....	450
Figure 8.1 : Les dimensions 1 et 3 de la régression PLS.....	483
Figure 8.2 : Les dimensions 2 et 4 de la régression PLS.....	488
Figure 8.3 : Corrélations entre variables prises deux à deux.....	495

LISTE DES TABLEAUX

Tableau 1.1 : La classification de diplôme retenue.....	69
Tableau 1.2 : La description des grandes écoles dans les dictionnaires des codes.....	71
Tableau 1.3 : Proportion des individus vivant en couple cohabitant (marié ou non) par âge et décennie d'enquête (en %)......	80
Tableau 1.4 : Répartition des individus par année d'enquête et cohorte de naissance (% en ligne).....	81
Tableau 1.5 : Notation des proportions des effectifs des différentes cellules dans une table d'homogamie.....	84
Tableau 1.6 : Table d'homogamie simplifiée dans une population peu diplômée.....	85
Tableau 1.7 : Table d'homogamie simplifiée dans une population relativement diplômée.....	86
Tableau 1.8 : Table d'homogamie simplifiée dans une population où les femmes sont plus diplômées que les hommes.....	87
Tableau 2.1 : Indicateurs d'ajustement des modèles concernant l'ensemble des configurations de couple.....	146
Tableau 2.2 : Indicateurs d'ajustement des modèles concernant les configurations exogames	156
Tableau 3.1 : Le codage des différentes définitions du chômage dans les enquêtes Emploi..	191
Tableau 3.2 : Le codage des emplois précaires dans les enquêtes Emploi.....	192
Tableau 3.3 : Indicateurs d'ajustement des modèles utilisés pour la décomposition.....	209
Tableau 3.4 : Indicateurs d'ajustement des modèles d'analyse de l'association.....	222
Tableau 3.5 : Indicateurs d'ajustement des modèles concernant la précarité de l'emploi....	232
Tableau 4.1 : Matrice d'hypergamie de classe sociale.....	253
Tableau 4.2 : Indicateurs d'ajustement des modèles.....	264
Tableau 4.3 : Indicateurs d'ajustement des modèles incluant les célibataires.....	280
Tableau 5.1 : Répartition des individus selon leur histoire conjugale et les données disponibles par cohorte et par sexe.....	295
Tableau 6.1 : Statut d'activité des conjoints lors de la première mise en couple d'Ego.....	349
Tableau 6.2 : Écart d'âge en faveur de l'homme selon le statut d'activité des deux conjoints	350
Tableau 6.3 : Statistiques d'ajustement des trois modèles.....	367
Tableau 6.4 : Résultats détaillés du modèle 1 appliqué séparément à deux générations.....	387
Tableau 6.5 : Résultats détaillés du modèle 2 appliqué séparément à deux générations.....	391
Tableau 7.1 : La classification d'éducation en trois niveaux de l'Enquête européenne sur les forces de travail.....	409
Tableau 7.2 : Matrice d'hypergamie socioéconomique.....	416
Tableau 7.3 : Indicateurs d'ajustement des modèles concernant l'homogamie.....	427
Tableau 7.4 : Indicateurs d'ajustement des modèles concernant l'hypergamie.....	447
Tableau 7.5 : Régions retenues et effectifs correspondants dans l'échantillon.....	454
Tableau 8.1 : Vagues de l'Enquête sociale européenne auxquelles ont participé les différents pays.....	463
Tableau 8.2 : Proportion de variance expliquée par les dimensions (R^2 en %)......	481
Tableau 8.3 : Valeur des variables et effectifs de l'Enquête sociale européenne pour chaque région étudiée.....	496

Tableau A.1 : Codage de la classe EGP, 1969-1974 (Code des métiers de 1968).....	518
Tableau A.2 : Codage de la classe EGP, 1975-1981 (Code des métiers de 1975).....	519
Tableau A.3 : Codage de la classe EGP, 1982-2002 (Code des PCS 1982).....	520
Tableau A.4 : Codage de la classe EGP, 2003-2011 (Code des PCS 2003).....	521
Tableau A.5 : Variables et modalités utilisées pour coder la classe EGP des conjoints.....	522
Tableau A.6 : Variables et modalités utilisées pour coder la classe EGP d'origine.....	523

LISTE DES ENCADRÉS

Encadré 1.1 : La nomenclature de classe Erikson-Goldthorpe-Portocarero.....	72
Encadré 1.2 : La représentation des tables d'homogamie en mosaïque.....	100
Encadré 2.1 : Équations des modèles mentionnés.....	136
Encadré 2.2 : Niveau 3 de la nomenclature des PCS 2003.....	174
Encadré 3.1 : Équations des modèles log-linéaires utilisés pour la décomposition.....	207
Encadré 3.2 : Équations des modèles log-linéaires utilisés pour l'analyse de l'association..	217
Encadré 4.1 : Équations des modèles.....	259
Encadré 4.2 : Équation du modèle d'association anti-symétrique.....	277
Encadré 5.1 : Équations et estimation de la distribution log-logistique avec immunité.....	305
Encadré 5.2 : Les splines cubiques restreintes.....	308
Encadré 6.1 : Le modèle de Cox.....	357
Encadré 7.1 : La classification socioéconomique européenne ESeG.....	414
Encadré 7.2 : Équations du modèle log-multiplicatif à forme de régression multiple.....	420
Encadré 8.1 : L'algorithme de la régression PLS.....	476

TABLE DES MATIÈRES

Sommaire.....	3
Remerciements.....	5
Introduction.....	11
1. Des évidences bien fragiles.....	17
1.1. Un maintien de l'homogamie ?.....	18
1.2. Une hypergamie inébranlable ?.....	19
2. L'invention de l'amour ?.....	20
2.1. Une « révolution sentimentale » surestimée.....	20
2.2. Des évolutions réelles.....	23
2.3. Les nouvelles formes du contrôle social.....	24
2.4. Les concepts de choix du conjoint et de marché conjugal.....	25
3. L'individualisation et le choix du conjoint chez les auteurs classiques.....	28
3.1. Tocqueville et l'effacement des lignées familiales.....	29
3.2. Durkheim et le développement de la « famille conjugale ».....	30
3.3. Simmel et l'entrecroisement des cercles sociaux.....	32
4. Les évolutions récentes des sociétés européennes.....	34
4.1. Apogée et déclin du modèle conjugal de l'après-guerre.....	34
4.2. L'affaiblissement du mariage et la montée des cohabitations.....	35
4.3. L'élévation du statut des femmes.....	37
4.4. L'évolution des modèles conjugaux et la « relation pure ».....	38
4.5. L'évolution de la stratification sociale.....	39
5. Le cadre d'analyse contemporain du choix du conjoint.....	42
5.1. La notion d'ouverture sociale.....	42
5.2. Théories de la modernisation et évolutions temporelles.....	44
5.3. Les quatre générations d'études de l'homogamie.....	46
6. Plan de la thèse.....	49
Partie I : Dimensions et évolutions de l'homogamie en France (1969-2011).....	53
Chapitre 1 : De la structure de la population à celle des couples : le déclin de l'homogamie absolue de diplôme, de classe et d'origine sociales.....	55
1. Des résultats en débat.....	56
1.1. La littérature internationale : des conclusions divergentes.....	56
1.2. Les études sur la France : résultats et limites.....	59
2. La construction d'une série harmonisée des enquêtes Emploi.....	63

2.1. Les sources disponibles.....	63
2.2. Les enquêtes Emploi.....	64
2.3. La définition des variables.....	67
2.3.1. Le diplôme.....	67
2.3.2. La classe EGP.....	72
2.3.3. La classe EGP d'origine.....	77
2.4. Le choix des limites d'âge et la définition des couples.....	77
3. De l'homogamie absolue à l'homogamie relative.....	81
3.1. La mesure absolue de l'homogamie : intérêt et limites.....	83
3.2. Le concept d'homogamie relative et sa mesure.....	88
4. De l'évolution de la structure de la population à celle des couples.....	92
4.1. Le diplôme : les effets de la massification scolaire.....	92
4.1.1. Une transformation majeure de la structure par diplôme de la population.....	92
4.1.2. Une nette diminution du taux d'endogamie de diplôme.....	94
4.1.3. Un effondrement de la proportion de couples peu diplômés.....	97
4.2. La classe sociale : déclin de l'indépendance et tertiarisation.....	102
4.2.1. Un déclin des indépendants et des ouvriers non qualifiés et une montée des cols blancs.....	102
4.2.2. Une nette diminution du taux d'endogamie de classe sociale.....	106
4.2.3. Un déclin des couples d'indépendants et une perte d'unité des classes populaires.....	108
4.2.4. La structure de l'homogamie vue à travers la PCS.....	114
4.3. L'origine sociale : une image en différé.....	114
4.3.1. Des origines moins souvent agricoles, et plus souvent en col blanc.....	114
4.3.2. Une diminution du taux d'endogamie d'origine sociale.....	115
4.3.3. Une lente évolution structurelle et un affaiblissement de l'homogamie.....	117
5. Conclusion.....	120

Chapitre 2 : L'homogamie relative de diplôme, de classe et d'origine sociales et ses évolutions : ouverture d'ensemble, repli des élites.....123

1. Modéliser l'homogamie relative.....	124
1.1. La sélection des modèles : les indicateurs BIC et AIC.....	126
1.2. Les limites de la sélection d'un modèle à partir d'un unique critère statistique.....	127
1.3. Les objectifs différents du BIC et de l'AIC.....	129
1.4. Des indicateurs absolus d'ajustement.....	133
1.5. La stratégie de modélisation.....	134
2. Mesurer l'intensité de l'homogamie relative.....	140
2.1. Les limites des indices existants.....	141
2.2. Le coefficient d'association intrinsèque λ de Goodman.....	142
2.3. L' <i>odds ratio</i> absolu moyen (MAOR).....	143
3. Un affaiblissement de l'homogamie relative.....	145

3.1. Une diminution de l'intensité générale de l'homogamie relative.....	145
3.2. Une baisse de l'endogamie relative de la plupart des groupes.....	149
4. Les transformations de l'espace social.....	154
4.1. Les modèles RC et RC-L.....	154
4.2. L'homogamie de diplôme : la fin de la spécificité des non diplômés.....	159
4.3. L'homogamie de classe sociale : une baisse générale de l'association.....	166
4.4. L'homogamie de classe sociale d'origine : une plus grande stabilité.....	175
5. Conclusion.....	178

Chapitre 3 : La situation sur le marché du travail : effet égalisateur de l'activité féminine mais stabilité du lien entre statuts d'emploi.....183

1. Les données : salaire et précarité de l'emploi dans les enquêtes Emploi.....	187
1.1. Le salaire.....	187
1.2. Les indicateurs de la précarité de l'emploi.....	189
2. La diminution des inégalités de salaire entre couples et ses composantes.....	193
2.1. Les conséquences de l'homogamie et de la hausse de l'activité féminine sur les inégalités.....	193
2.2. Les trois composantes de l'inégalité entre salaires des couples.....	196
2.3. La distribution sociale de l'inactivité.....	201
3. Les facteurs de l'évolution des inégalités : scénarios contrefactuels et évolutions localisées.....	203
3.1. La méthode de décomposition par la modélisation log-linéaire.....	204
3.2. Résultats des modèles : les contributions aux inégalités.....	208
4. L'évolution de l'association entre salaires des conjoints et de l'hypergamie au sein des couples bi-actifs.....	213
4.1. La spécification de l'association entre salaires des conjoints.....	214
4.1.1. L'association symétrique.....	216
4.1.2. L'hypergamie.....	216
4.2. Résultats : affaiblissement de l'association, résistance de l'hypergamie.....	220
4.2.1. Ajustement des modèles.....	220
4.2.2. Un affaiblissement de l'association en termes positionnels et monétaires.....	223
5. Une stabilité de l'association entre statuts d'emploi des conjoints.....	228
5.1. Une explosion de la proportion de couples formés de deux précaires.....	229
5.2. Une stabilité de l'endogamie relative du groupe des précaires.....	231
6. Conclusion.....	234

Partie II : Mécanismes de l'homogamie et de l'hypergamie en France.....237

Chapitre 4 : La disparition de l'hypergamie relative de diplôme, de classe et d'origine sociales : une convergence du célibat entre les sexes (1969-2011) 239

1. Perspectives théoriques et état des connaissances.....	241
---	-----

1.1. La conception classique de l'hypergamie féminine : un modèle fortement généré.....	241
1.2. L'évolution temporelle de l'hypergamie : une transformation majeure mais incomplète ?.....	245
1.3. Les travaux existants : baisse et inversion de l'hypergamie de diplôme.....	249
2. L'hypergamie absolue : entre affaiblissement et inversion.....	251
2.1. La définition de l'hypergamie : hiérarchiser les catégories.....	251
2.2. L'inversion de l'hypergamie absolue de diplôme.....	254
2.3. Le développement de l'hypogamie féminine de classe sociale.....	255
2.4. L'absence d'hypergamie selon la classe sociale d'origine.....	257
3. La fin de l'hypergamie relative.....	258
3.1.1. L'inversion de l'hypergamie relative de diplôme.....	263
3.1.2. La quasi-disparition de l'hypergamie relative de classe sociale.....	266
3.1.3. L'introuvable hypergamie relative d'origine sociale.....	267
4. L'évolution du célibat : une convergence entre hommes et femmes.....	268
4.1. Une convergence entre hommes et femmes du célibat selon le diplôme.....	269
4.2. Des différences générées moindres mais persistantes selon la classe sociale	272
4.3. Une atténuation des différences selon la classe sociale d'origine chez les femmes.....	273
5. Une synthèse : le modèle d'association anti-symétrique.....	274
5.1. Le modèle d'association anti-symétrique.....	275
5.2. L'inversion des asymétries selon le diplôme.....	278
5.3. La disparition des asymétries selon la classe sociale.....	282
5.4. L'affaiblissement des asymétries selon la classe sociale d'origine.....	286
6. Conclusion.....	288

Chapitre 5 : Les mécanismes du célibat définitif des diplômées : la fin de la Sainte-Catherine (cohortes 1920-1970).....291

1. Les données : Étude de l'histoire familiale et Biographies et entourage.....	293
2. L'inversion de l'hypergamie de diplôme au fil des cohortes.....	297
2.1. L'hypogamie féminine est devenue la situation la plus fréquente.....	298
2.2. L'inversion persiste au-delà des évolutions structurelles.....	302
3. Modéliser la première mise en couple.....	304
3.1. Le modèle log-logistique avec immunité.....	304
3.2. L'adéquation du modèle aux données.....	310
4. L'évolution et les mécanismes du célibat définitif.....	311
4.1. La résorption du surcroît de célibat des diplômées.....	311
4.2. Le renforcement du lien entre diplôme et âge de mise en couple chez les hommes.....	315
4.3. L'allongement de la période de mise en couple des femmes.....	318
5. Conclusion.....	321

Chapitre 6 : Les déterminants biographiques de l'homogamie et de l'hypergamie de diplôme au sein des premiers couples : l'âge et la scolarisation (cohortes 1920-1970).....	325
1. Contexte biographique de la mise en couple et choix du conjoint : théorie et littérature existante.....	328
1.1. L'âge de mise en couple.....	328
1.1.1. Âge et contraintes structurelles sur le marché conjugal.....	328
1.1.2. Des effets différents de l'âge selon le sexe ?.....	330
1.2. La scolarisation.....	333
1.2.1. L'influence des études sur les fréquentations.....	333
1.2.2. Des fréquentations scolaires aux rencontres.....	335
1.2.3. Les effets de la sortie du système scolaire.....	337
1.3. Les travaux existants.....	338
2. Mise en couple et scolarisation dans les enquêtes Étude de l'histoire familiale et Biographies et entourage.....	339
2.1. Étude de l'histoire familiale.....	339
2.2. Biographies et entourage.....	341
3. Mise en couple et scolarisation : un état des lieux.....	343
3.1. Le développement des mises en couple pendant et peu après les études...	343
3.2. Le rapprochement entre fin d'études et mise en couple.....	346
3.3. Qui sont les couples formés en cours d'études ?.....	348
4. Modéliser les chances de mise en couple : méthode et variables.....	353
4.1. Le modèle de Cox à risques concurrents.....	354
4.2. Variables.....	359
4.2.1. La mesure du niveau d'éducation : causalité et anticipation.....	359
4.2.2. Origine sociale.....	361
4.2.3. Variables biographiques d'intérêt : temps et scolarisation.....	362
4.2.4. Année de naissance.....	363
4.3. Spécification des modèles.....	363
5. Résultats des modèles à risques concurrents.....	365
5.1. Validation des modèles.....	365
5.2. L'évolution au cours du temps du niveau de diplôme des conjoints : une convergence entre hommes et femmes.....	367
5.2.1. L'hypergamie augmente avec l'âge.....	367
5.2.2. L'âge de mise en couple n'affecte pas significativement l'endogamie	373
5.2.3. Années depuis la fin des études : une confirmation de l'effet de l'âge	373
5.3. La scolarisation augmente l'hypergamie et l'endogamie.....	375
5.3.1. La scolarisation augmente l'hypergamie.....	375
5.3.2. La scolarisation augmente l'endogamie.....	377
5.4. La durée des études : une confirmation partielle de l'effet de la scolarisation	379
5.4.1. Le temps de scolarisation augmente l'hypergamie.....	380
5.4.2. Le temps de scolarisation n'augmente pas l'endogamie.....	381

5.5. Le temps écoulé depuis la fin des études diminue l'endogamie.....	382
6. Conclusion.....	383
7. Appendice : Résultats détaillés des modèles à risques concurrents estimés séparément sur deux cohortes.....	385
Partie III : Variations et déterminants de l'homogamie et de l'hypergamie en Europe.....	395
Chapitre 7 : La structure et le niveau de l'homogamie et de l'hypergamie socioéconomique et d'éducation : des différences marquées entre et à l'intérieur des pays.....	397
1. Littérature : des débats centrés sur les effets du développement et de l'ouverture sociale.....	399
1.1. L'hypothèse d'une relation en U inversé avec le développement.....	400
1.2. Les études récentes : des données harmonisées.....	402
1.3. Une synthèse possible ?.....	404
2. Données : l'Enquête européenne sur les forces de travail.....	406
2.1. Choix et définition des unités géographiques : les régions NUTS1.....	406
2.2. Variables : niveau d'éducation et groupe socioéconomique.....	408
2.2.1. Le niveau d'éducation.....	408
2.2.2. Le groupe socioéconomique.....	412
2.2.3. La définition de l'hypergamie.....	415
2.3. Champ : les jeunes couples cohabitants.....	416
3. La stratégie de modélisation : le modèle à forme de régression multiple.....	417
3.1. Modéliser les variations de l'homogamie.....	417
3.2. Modéliser les variations de l'hypergamie.....	422
4. Les variations de l'homogamie.....	423
4.1. L'homogamie absolue : une correspondance entre composantes éducative et socioéconomique.....	423
4.2. L'homogamie relative : portrait de huit régions-types.....	426
4.2.1. La géographie de l'homogamie relative en Europe.....	428
4.2.2. La structure de l'homogamie d'éducation : l'influence de l'organisation du système scolaire.....	431
4.2.3. La structure de l'homogamie socioéconomique.....	438
5. Les variations de l'hypergamie.....	442
5.1. L'hypergamie absolue : plus fréquente en termes socioéconomiques qu'éducatifs.....	442
5.2. L'hypergamie relative : une tendance loin d'être systématique.....	445
5.2.1. La significativité statistique de l'hypergamie relative.....	447
5.2.2. La géographie de l'hypergamie relative en Europe.....	449
6. Conclusion.....	452
7. Appendice : Détail des régions étudiées.....	454

Chapitre 8 : Les facteurs de l'homogamie et de l'hypergamie socioéconomique et d'éducation : ouverture sociale et libéralisme culturel plus que niveau de développement.....	459
1. Données : l'Enquête sociale européenne et d'autres sources.....	461
2. Les déterminants de l'homogamie : des hypothèses aux variables.....	463
2.1. Opportunités de rencontre.....	463
2.2. Entrecroisement des appartenances.....	465
2.3. Niveau de vie.....	466
2.4. Inégalités sociales et risque de mésalliance.....	467
2.5. Libéralisme des mœurs et contrôle social.....	468
2.6. Religion.....	471
3. Le statut des femmes et les déterminants de l'hypergamie féminine.....	472
3.1. Activité féminine et choix du conjoint : un lien débattu.....	472
3.2. Statut des femmes et hypergamie.....	473
4. Méthode : la régression PLS.....	475
4.1. Principe : une analyse factorielle tournée vers l'explication.....	475
4.2. Du « problème du petit N » à une analyse « avec noms propres ».....	478
4.3. La modélisation de plusieurs variables.....	479
4.4. Estimation.....	480
5. Résultats : des déterminants largement communs.....	480
5.1. Les déterminants communs de l'homogamie et de l'hypergamie : libéralisme culturel, ouverture sociale et niveau de développement.....	482
5.2. Des divergences : des aires géographiques aux structures sociales contrastées.....	487
6. Conclusion.....	491
7. Appendice : Valeurs des variables et corrélations deux à deux.....	495
Conclusion.....	503
Annexes.....	515
Annexe A : Le codage de la classe Erikson-Goldthorpe-Portocarero (EGP) dans les enquêtes Emploi.....	517
Annexe B : Le coefficient d'association intrinsèque et l'<i>odds ratio</i> absolu moyen (MAOR), deux mesures de l'association indépendantes des marges.....	525
1. Un cadre général réunissant tradition pearsonienne et <i>odds ratios</i>	527
1.1. Définitions préliminaires.....	527
1.2. Le coefficient de contingence carrée moyenne de Pearson.....	530
1.3. Le coefficient d'association intrinsèque.....	531
2. L' <i>odds ratio</i> absolu moyen (MAOR).....	535

2.1. MAOR avec pondération uniforme.....	536
2.2. MAOR avec pondération arbitraire.....	540
3. Relation avec le modèle UNIDIFF.....	545
4. Relation avec les modèles d'association lignes-colonnes.....	547
4.1. Le coefficient d'association intrinsèque.....	548
4.2. Le MAOR dans les modèles RC(M).....	550
4.3. Extension au modèle RC(M)-L.....	550
5. Composantes symétrique et anti-symétrique de l'association.....	552
5.1. Le coefficient d'association intrinsèque.....	553
5.2. Le MAOR symétrique et anti-symétrique.....	555
5.3. Le modèle d'association lignes-colonnes symétrique.....	557
5.4. Le modèle d'association lignes-colonnes anti-symétrique.....	558
Bibliographie.....	563
Liste des figures.....	595
Liste des tableaux.....	599
Liste des encadrés.....	601