

ANALYSE DES SÉQUENCES ET OPTIMAL MATCHING : LE PASSAGE À L'ÂGE ADULTE DES FEMMES ET DES HOMMES EN FRANCE

Elisabeth MORAND (*), Laurent TOULEMON (**)

(*) Ined, Service des Méthodes Statistiques

(**) Ined, unité Fécondité, famille, sexualité

Introduction

Dans un précédent travail, Winkler-Dworak et Toulemon (2007) ont comparé les modalités du passage à l'âge adulte à partir de modèles de durée, décrivant les évolutions des mises en couple et des premières naissances et leurs variations selon la situation face à la fin des études et du premier emploi. Les contraintes auxquelles sont soumis les femmes et les hommes en termes de mise en couple et de naissance du premier enfant ont eu tendance à se rapprocher, entre les générations nées au début des années 1940 et celles nées au début des années 1960.

Dans le cadre d'une approche globale de la transition vers l'âge adulte (Galland 2000, Galland 2007, Robette 2008), cet article cherche à analyser les transitions vers l'âge adulte en les décrivant de manière globale, plutôt qu'en considérant *a priori* que certains événements influent sur la survenue d'autres événements.

Reprenant la question des différences entre hommes et femmes et de leur évolution, nous décrivons le passage à l'âge adulte comme une succession de passages : fin des études, premier emploi, première union, premier mariage, naissance du premier enfant. D'une part nous considérons ces passages comme irréversibles (par exemple une personne exerçant pour la première fois une activité professionnelle passe de l'état « n'a jamais travaillé » à l'état « a déjà travaillé ». Si elle interrompt ensuite son activité, elle ne change pas d'état pour nous. Nous décrivons donc des « états » définis par les étapes qui ont été franchies pour y parvenir, plus que par la situation concrète à laquelle ils correspondent. Comme les passages sont très corrélés entre eux, nous avons choisi de ne pas prendre en compte le départ de chez les parents, qui survient presque toujours avant la mise en couple.

Après une rapide présentation des données et des méthodes d'*optimal matching* permettant de calculer des distances entre des séquences d'états, nous décrivons la diversité des situations des jeunes adultes. Puis nous décrivons la complexité des séquences à l'aide des mesures de l'entropie et de la turbulence, en précisant la différence entre ces deux indicateurs. Enfin, nous reviendrons sur la question de l'évolution de la différence entre femmes et hommes pour comparer nos résultats avec ceux obtenus par des méthodes radicalement différentes

1. Données et méthodes

1.1. Données

Les données utilisées sont celles de l'enquête « Étude de l'Histoire Familiale » réalisée en 1999 dans le cadre du recensement général de la population (Cassan et al. 2000 ; Lefèvre et Filhon 2005). Quelques 235 000 femmes et 145 000 hommes ont rempli un questionnaire de quatre pages sur leurs enfants, les unions, leur vie professionnelle, leurs origines familiales, les langues reçues de leurs parents et transmises à leurs enfants. Nous utilisons le sous-échantillon des hommes et des femmes

nés entre 1940 et 1964, pour décrire le passage à l'âge adulte de générations qui ont eu 20 à 30 ans dans les années 1960 à 1999. Nous avons utilisé l'échantillon de travail de Winkler-Dworak et Toulemon (2007) : les immigrants arrivés après l'âge de 15 ans ont été exclus, pour que tous les événements que nous décrivons entre les âges 15 et 35 dans ces générations aient eu lieu en France. nous avons également exclu les (très rares) cas où une union ou une naissance a été déclarée avant l'âge de 15 ans. L'échantillon retenu regroupe 140 000 femmes et 95 000 hommes. Les dates des événements suivants sont recueillies : âge à la fin des études initiales, âge au premier emploi (de plus de trois mois), date (mois et année) de la première union (de plus de six mois, les unions plus courtes n'étant pas décrites dans l'enquête), du premier mariage et de la naissance du premier enfant. On observe la situation de chaque personne à chaque anniversaire entre 15 et 35 ans compris ; pour la naissance de l'enfant on considère la date de conception et non la date de la naissance, pour mieux observer les comportements simultanés. On trouvera une définition précise des questions utilisées dans le questionnaire et du codage des variables dans Winkler-Dworak et Toulemon (2007). On caractérise les personnes par leur sexe et leur date de naissance (regroupés en générations de 5 années). La génération la plus ancienne (notée 1940) est celle des individus nés entre 1940 et 1944 inclus et la génération la plus récente (notée 1960) comprend les personnes nées entre 1960 et 1964 inclus.

Pour chaque individu, on s'intéresse donc à cinq événements (ou passages) constitutifs du passage à l'âge adulte : la fin des études (S pour *Studies*), le premier emploi (J pour *Job*), la première mise en couple (U pour *Union*), le premier mariage (M pour *Marriage*¹) et le premier enfant (C pour *Conception of a Child*)². À chaque anniversaire entre 15 et 35 ans, on observe lesquels de ces 5 événements se sont déjà produits pour la personne.

En observant la situation à chaque anniversaire, on sépare deux événements proches dans le temps mais ayant lieu à des âges différents (par exemple 20 ans 11 mois et 21 ans 1 mois), tandis que des événements moins proches (21 ans 1 mois et 21 ans 11 mois) sont considérés comme simultanés. L'absence d'information sur les mois de fin d'études et de premier emploi nous impose cette contrainte.

1.2. Notations

Pour un individu, une séquence est l'ensemble des 21 états (ou situations) observés à chaque âge entre 15 et 35 ans. L'état à un âge donné résulte des événements (ou étapes) qui se sont produits au cours de la période précédent l'anniversaire de l'individu. Ainsi un individu est dans l'état SJ à 19 ans s'il a terminé ses études et obtenu un premier emploi avant son 19^{ème} anniversaire. Ce même individu sera dans l'état SJC à 20 ans si entre 19 et 20 ans il a eu un premier enfant. Au moment de fêter ses 20 ans, cet individu aura fini ses études, obtenu un premier emploi et eu un premier enfant.

A chaque âge entre 15 et 35 ans, chaque individu est dans un état pris parmi 24 combinaisons possibles³. Des événements ayant lieu au même âge sont considérés comme simultanés.

Deux conventions sont choisies pour définir ces états. Premièrement, les passages sont irréversibles, et les passages d'un état à l'autre ne peuvent se faire que par le franchissement d'une ou plusieurs étapes supplémentaires. Ainsi, l'état SJ ne peut mener qu'à SJU, SJC, SJUM, SJUMC (=A). Deuxièmement, la situation à un âge donné ne dépend pas de la séquence des étapes qui ont déjà été franchies : on note SJ la situation d'une personne ayant fini ses études et exercé un emploi, sans distinguer laquelle de ces deux étapes, S ou J, a été franchie la première. Par contre les séquences diffèrent si les événements n'ont pas eu lieu au même âge : si un individu a fini ses études à 15 ans et travaillé à 16 ans, la séquence de ses états à 15, 16 et 17 ans sera N-S-SJ ; s'il a d'abord travaillé à 15 ans puis fini ses études à 16 ans, la séquence de ses états à 15, 16 et 17 ans sera N-J-SJ ; les états aux 15 et 17 anniversaires sont identiques, seul change l'état à 16 ans.

¹ Par convention on suppose que les mariés vivent ensemble : U est donc toujours simultané ou antérieur à M.

² Nous datons la naissance par la date de conception, de manière à considérer, par exemple, une conception pré-nuptiale comme une conception avant mariage, et non comme une naissance dans le mariage. Seules les conceptions ayant donné lieu à une naissance vivante sont prises en compte.

³ Avec 5 événements qui peuvent s'être produits ou non, 32 états sont possibles ($2^5=32$). Comme le mariage implique l'union, les états contenant M et non U sont impossibles ($2^3=8$ états impossibles). Restent donc 24 états possibles.

1.2.1. Séquences répétées

La biographie d'un individu est décrite par la séquence constituée des 21 états pris à chaque anniversaire entre 15 et 35 ans. Chaque individu durant la période peut passer au maximum par 6 états distincts par convention car aucun état n'est réversible, et les étapes s'ajoutent les unes aux autres. Les états ne sont donc pas indépendants d'un âge à l'autre : un état correspondant à certains passages déjà franchis ne peut mener qu'aux états qui intègrent ces passages, auxquels s'ajoutent un ou plusieurs autres.

Une séquence individuelle juxtapose les états aux anniversaires 15 à 35 ; elle est donc de la forme :

N-N-J-J-J-J-J-SJ-SJUM-SJUM-SJUM-A-A-A-A-A-A-A-A

Ici l'individu a exercé son premier emploi à 16 ans, puis a terminé ses études à 21 ans, il s'est marié l'année suivante et a eu un premier enfant 3 ans plus tard. On considère, s'il n'y a pas eu de première union déclarée avant le premier mariage, que le premier mariage est une première union.

Ainsi tant qu'un nouvel évènement ne s'est pas produit, l'individu reste dans le même état. Les durées passées dans chaque état, mesurées par le nombre de fois où l'individu est repéré dans un état, différencient les séquences qui partagent la même suite d'états.

1.2.2. Séquences non répétées

Si l'on ne tient pas compte du nombre d'années passées entre deux évènements, on définit des séquences non répétées pour résumer le parcours des individus à partir de la suite d'étapes. Ainsi la séquence non répétée correspondant à l'exemple précédent est N-J-SJ-SJUM-A. Les séquences non répétées décrivent les étapes franchies et l'ordre dans lesquels elles ont été franchies. Une même séquence non répétée correspond à plusieurs parcours décrits par des séquences répétées différentes.

A la différence des séquences répétées qui sont toutes de longueur 21 éléments, les séquences non répétées ont des longueurs variant de 1 à 6 éléments.

1.3. Méthodes

La distance entre les séquences est obtenue par optimal matching. Cette distance est ensuite utilisée pour regrouper les séquences en classes par une classification ascendante hiérarchique.

1.3.1. L'optimal matching

La méthode utilisée ici pour comparer entre elles les séquences (répétées ou non répétées) est celle de l'« optimal matching ». Cet algorithme de comparaison a été en premier lieu utilisé pour l'étude des séquences d'ADN ou de protéines. Les algorithmes d'optimal matching en sciences sociales ont été d'abord utilisés sur des données historiques (Abbott et Forest, 1986).

L'objectif de ces méthodes est de calculer une distance entre les séquences prises deux à deux. La distance entre deux séquences dépend du nombre d'opérations nécessaires pour transformer une séquence en l'autre. Il existe deux types d'opérations : insérer ou supprimer un état d'une part, substituer un état par un autre d'autre part. On associe un coût à chacune de ces opérations.

Un choix classique de coût de un pour l'insertion-suppression et de deux pour la substitution tend à réduire le temps à la seule dimension du classement (Kruskal 1983, Lesnard et Saint Pol 2006). Il est possible d'éviter les déformations du temps en favorisant les opérations de substitutions par rapport aux opérations d'insertions suppressions. Les coûts de substitutions doivent alors être faibles. On peut aussi proposer des coûts de substitutions différents pour chaque état à l'aide des taux de transitions observés sur les données.

Pour les analyses effectuées ici on a choisi de donner un coût de 1 à l'insertion et à la suppression et un coût de 2 à la substitution, qui peut se réaliser par un ajout suivi d'une suppression Ceci revient à

utiliser comme distance entre les séquences la distance LCS (*longest common subsequence*) décrite par Elzinga (2006). Pour les séquences répétées cette distance est proche de celle que l'on obtient à partir d'une analyse des correspondances multiples, car la distance compte le nombre d'éléments de la séquence qui diffèrent. La distance est cependant différente dans le cas de séquences décalées. Par exemple les séquences

N-N-**J**-J-J-J-**SJ-SJUM**-SJUM-SJUM- **A** -A-A-A-A-A-A-A-A
et N-N-**N**-J-J-J-J- **J** - **SJ** -SJUM-SJUM-**SJUM**-A-A-A-A-A-A-A-A

diffèrent en 4 âges (indiqués en italiques gras) ; quatre substitutions sont nécessaires pour passer de la première à la seconde. Mais il suffit d'insérer un N en début de séquence et de retirer un A en fin de séquence pour passer de la première à la seconde. La distance entre ces deux séquences serait donc de 4 pour la distance de l'ACM (nombre de paires discordantes entre les deux séries) et de 2 pour l'optimal matching (une insertion -N- et une délétion -A-). D'autres solutions sont envisageables comme les substitutions successives mais ces solutions couteraient toutes "plus chères" c'est donc ce coût minimum de 2 qui sera pris comme distance entre les deux séquences.

1.3.2. La classification ascendante hiérarchique

L'optimal matching permet d'obtenir une distance entre chaque séquence deux à deux. L'objectif est ensuite de positionner ces séquences dans des groupes homogènes (les séquences d'un même groupe sont proches au sens de la distance obtenue par optimal matching) et distincts (des séquences dans deux groupes différents se ressemblent moins que des séquences dans un même groupe). On utilise pour créer ces groupes des algorithmes de classification. Nous avons ici choisi d'utiliser une classification ascendante hiérarchique (CAH) utilisant l'algorithme de WARD⁴. Cette technique se fonde sur l'agrégation progressive des individus ou des groupes d'individus permettant à chaque étape de limiter au maximum la perte d'information due au regroupement ; ce critère revient à minimiser la variance intra-groupe et à maximiser la variance inter-groupes.

1.3.3. Les logiciels utilisés

La construction des variables et la constitution des fichiers d'analyse ont été réalisées avec le logiciel SAS v9. Les statistiques descriptives sur les séquences ainsi que les calculs de distances par optimal matching ont été faites avec la package *Traminer* du logiciel R (Gabadinho et al., 2008)

2. Description des parcours de passage à l'âge adulte de 15 à 35 ans

Dans un premier temps, on décrit les parcours des individus étudiés d'un point de vue global en considérant, à chaque âge, la répartition des individus entre les états possibles. Puis on effectue une description individuelle des trajectoires.

2.1. Description des états à chaque âge

2.1.1. La répartition des états à chaque âge

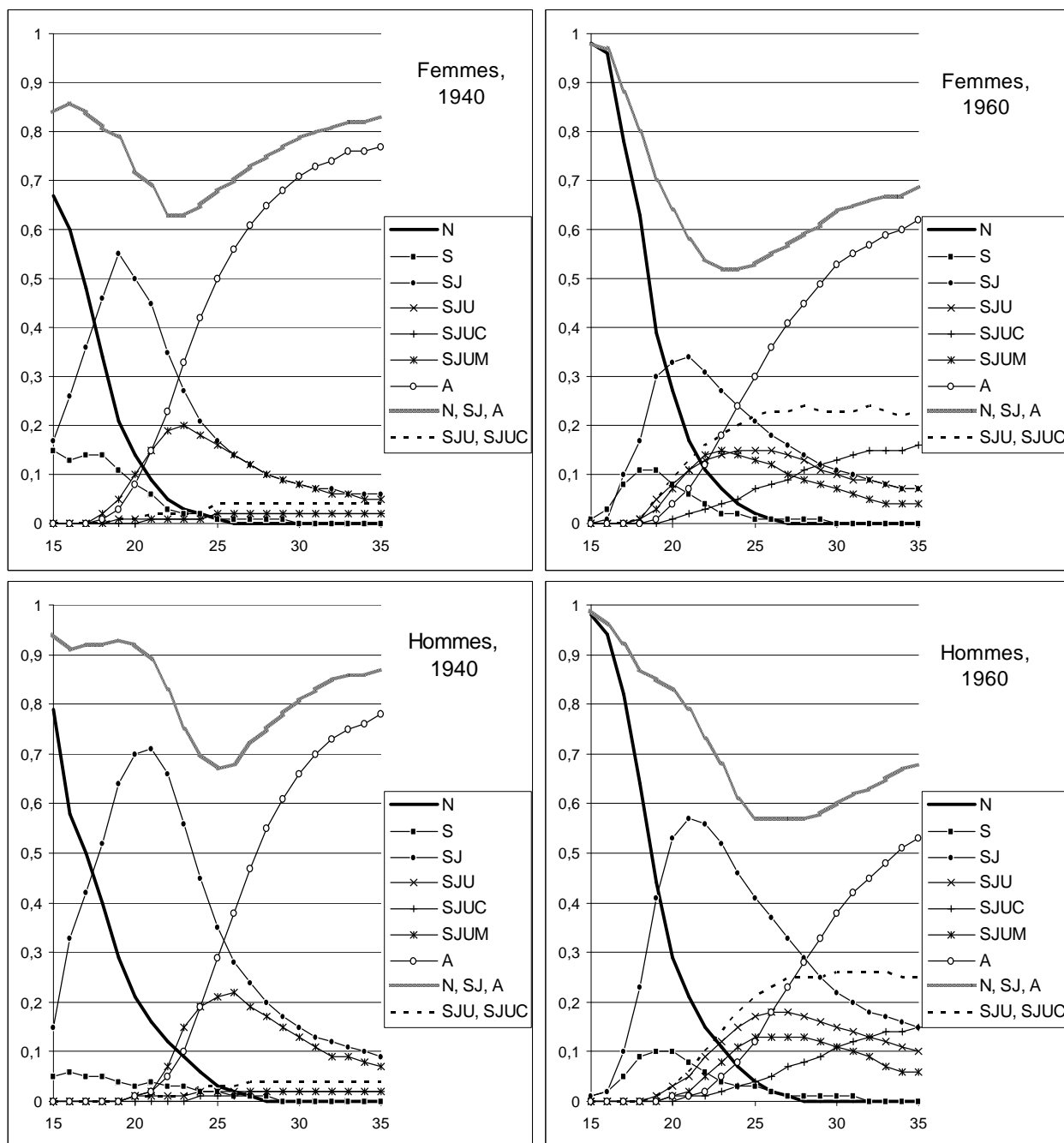
La figure 1 présente, pour les femmes et les hommes des générations nées en 1940-44 et 1960-64, la proportion de chaque état à chaque âge. Seuls les 7 états pris par au moins 3% de la population à au moins un âge donné se trouvent sur la figure 1 : N, S, SJ, SJU, SJUC, SJUM et A. On a rajouté la part

⁴ Ce dernier est peu approprié pour le cas où la distance employée n'est pas euclidienne (Saporta 2006) ce qui est en général le cas des distances obtenues par optimal matching (Lesnard et de Saint Pol 2006). Pour nos données, la matrice de distances obtenue est cependant très proche d'une matrice de distance euclidienne, ce qui justifie l'utilisation de cet algorithme. La matrice des produits scalaires déduite de la matrice de distance par la formule de Torgerson est en effet semi-définie positive, ce qui caractérise une matrice de distance euclidienne (Saporta 2006, Schoenberg J, 1937).

de ceux qui sont dans un des trois états les plus fréquents : N, SJ et A, ainsi que ceux qui vivent en couple sans être mariés, après avoir terminé leurs études et avoir travaillé (SJU, SJUC), deux états dont la fréquence devient importante dans la génération 1960.

Certains états sont beaucoup plus fréquents que d'autres : N aux jeunes âges, puis SJ aux âges moyens, quand les individus ont fini leurs études et exercé un premier emploi, puis A quand ils se sont mis en couple, se sont mariés et ont eu un enfant. Ces trois états regroupent à chaque âge au moins 63% des femmes et 67% des hommes nés en 1940, en moyenne femmes et hommes sont dans un de ces trois états respectivement 16 et 17 fois entre 15 et 35 ans (courbes gris clair sur la figure 1). Les mises en couple étant plus tardives pour les hommes, l'état A est davantage concentré aux âges élevés pour eux. Dans la génération 1960, les mêmes états dominent la répartition, mais moins fortement (au moins 52% pour les femmes et 57% à chaque âge, 14 et 15 fois respectivement en moyenne entre 15 et 35 ans). L'état N est plus fréquent à 15 ans et plus durable, en raison du retard de la fin des études et du premier emploi. Le premier emploi étant un peu plus retardé que la fin des études, l'état S devient plus tardif, pour les hommes comme pour les femmes et un peu plus fréquent pour les hommes, mais par pour les femmes, qui travaillent plus souvent dans la génération 1960 que dans la génération 1940. L'état A, très fréquent dans la génération 1940 (77% des femmes et 78% des hommes à 35 ans) diminue beaucoup : dans la génération 1960 ce ne sont plus que 62% des femmes et 53% des hommes qui ont connu toutes les étapes. Les états SJU et SJUC apparaissent dans la génération 1960, 16% des états de l'ensemble, les fréquences dépassant 20% après 25 ans, contre 2% maximum pour la génération 1940, courbes en traits pointillés sur la figure 1), en raison de la désaffection des jeunes couples pour le mariage (Toulemon 1996 ; Robette 2008). Globalement les situations semblent plus diverses pour la génération 1960 que pour la génération 1940.

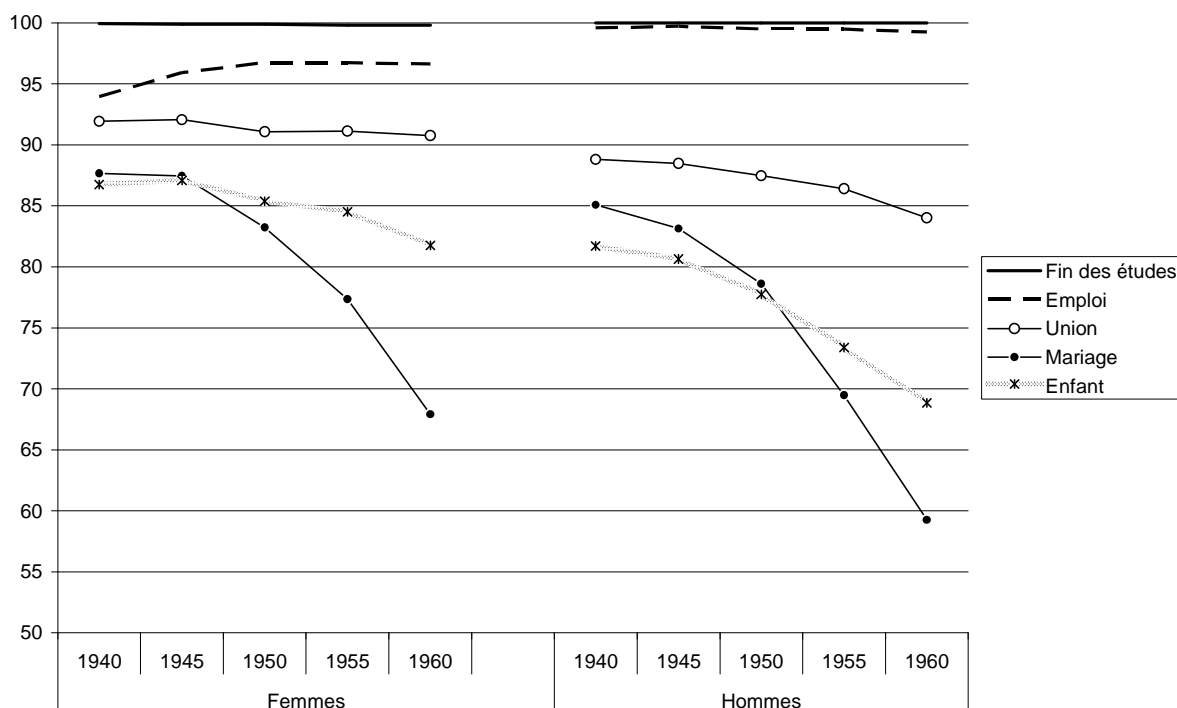
Figure 1. Répartition des individus entre les différents états selon l'âge, de 15 à 35 ans. Femmes et hommes nés en 1940-44 et 1960-64



2.1.2. Nombre d'états et d'étapes dans les séquences

La diversité des situations aux différents âges dépend de deux facteurs. Tout d'abord, le nombre moyen d'événements vécus avant 35 ans diminue au fil des générations. La proportion de femmes ayant exercé un emploi augmente au fil du temps, mais elle est déjà très élevée pour les femmes nées en 1940-44. Pour tous les autres événements, la fréquence avant 35 ans diminue, légèrement pour les mises en couple et les premières naissances, plus fortement pour les mariages (figure 2). La diminution apparaît plus marquée pour les hommes que pour les femmes : l'état SJ, caractéristique de ceux qui ne se sont pas mis en union, est le deuxième état le plus fréquent pour les hommes nés en 1960-64, derrière A (voir Figure 1), tandis que pour les femmes l'état SJUC est plus fréquent.

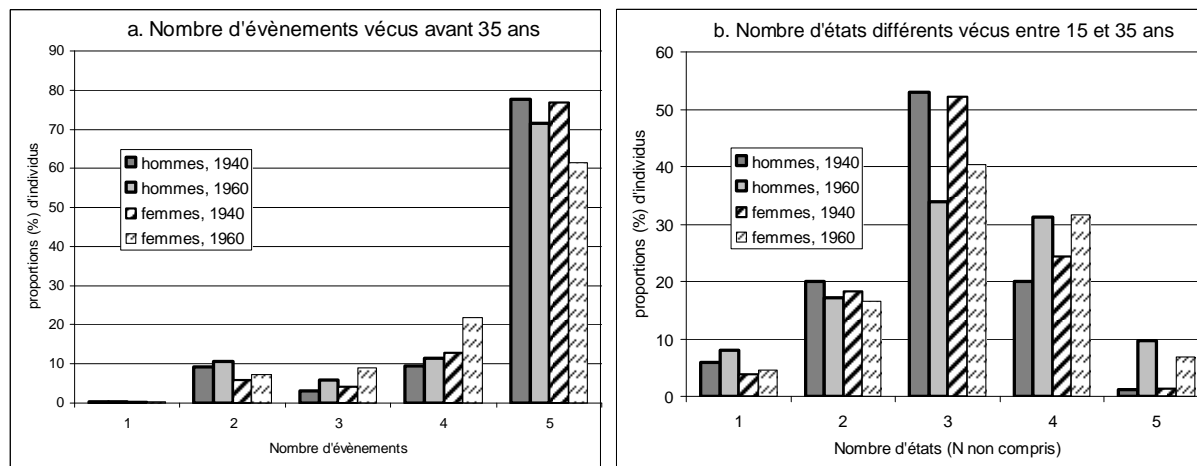
Figure 2. Proportions d'hommes et de femmes ayant vécu les différents événements avant 35 ans, selon la génération de naissance



On pourrait croire que plus le nombre d'événements vécus par les jeunes adultes entre 15 et 35 ans est élevé, plus les séquences peuvent être diversifiées. Comme le nombre moyen d'événements diminue, cela conduirait à une baisse au fil des générations du nombre d'états constituant les séquences de la transition vers l'âge adulte. Mais il n'en est rien : la désaffection des jeunes pour le mariage conduit à une diminution du nombre moyen d'événements, la proportion de ceux et celles qui ont connu les cinq événements avant 35 ans diminuant, de 77% à 62% pour les femmes, de 78% à 72% pour les hommes, tandis qu'augmentent la part de celles et ceux qui vivent 2, 3 ou 4 événements (Figure 3.a). Mais les événements sont de moins en moins souvent simultanés (au sens d'être vécus au même âge), et le nombre d'états différents vécus à partir du premier événement augmente : parmi les femmes et les hommes nés en 1940-44, presque aucun (environ 1%) n'était « passé » par 5 états différents ; dans la génération 1960-64, ce sont 7% des femmes et 10% des hommes qui ont vécu les différents événements à des âges tous différents ; les proportions de ceux qui ont expérimentés 4 états différents augmentent également, au détriment de ceux qui sont passés par 3 états, situation la plus fréquente dans la génération 1940 (Figure 3.b). En termes de séquences d'états non répétés, la séquence N-SJ-SJUM-A, décrivant trois changements : fin des études et premier emploi au même âge, union et mariage simultanés, puis conception d'un enfant, passe d'une fréquence de 25% dans la génération 1940, les deux sexes confondus, à 11%, tandis que la séquence similaire, mais avec la mise en couple et le mariage survenant à des âges différents, N-SJ-SJU-SJUM-A, passe de 1% à 8% ; on observe la même tendance pour la séquence N-S-SJ-SJUM-A (proche de la précédente, avec un premier emploi après la fin des études), remplacée par N-S-SJ-SJU-SJUM-A, par l'apparition de l'état de cohabitation sans mariage. La désaffection du mariage n'implique pas nécessairement une

diminution du nombre de situations différentes, la diffusion des séquences N-SJ-SJU-SJUC et N-S-SJ-SJU-SJUC se faisant au détriment de N-SJ-SJUM-A et N-S-SJ-SJUM-A, à nombre de situations différentes constant, la mise en couple se faisant sans mariage et non plus avec un mariage simultané.

Figure 3. Répartition des femmes et des hommes nés en 1940-44 et 1960-64 selon le nombre d'évènements vécus avant 35 ans, et le nombre d'états différents vécus entre 15 et 35 ans.



2.1.3. Nombre d'états et d'étapes à chaque âge

Nous avons vu que, bien que le nombre moyen d'évènements vécus diminue, le nombre différents d'états augmente. Le calendrier des transitions se modifie également : le nombre d'états différents vécus avant 20 ans reste constant chez les hommes, et diminue chez les femmes, tandis que le nombre d'états vécus augmente après 30 ans pour les hommes et après 25 ans pour les femmes (Figure 4).

Retard des transitions, diminution du nombre d'évènements et séparation croissante des évènements se compensent entre 20 et 25 ans pour les femmes, entre 20 et 30 ans pour les hommes. C'est que le nombre d'états vécus résulte de différents évènements qui suivent des logiques temporelles spécifiques. Ainsi l'âge moyen de début d'union est resté pratiquement stable, tandis que la fin des études et le premier emploi, d'une part, la naissance du premier enfant et, surtout, le mariage, d'autre part, devenaient plus tardifs (Figure 5). Dans la génération 1940-44, la fin des études est fortement dissociée des autres évènements, qui sont tous assez proches. Pour les hommes, le premier emploi est très proche de la fin des études ; pour les femmes, l'âge moyen doit être interprété avec prudence, car les âges au premier emploi sont fortement dispersés. Les âges moyens sont répartis en trois groupes : fin des études et premier emploi, mise en couple et mariage (qui se dissocient dans les générations 1960-64) et naissance du premier enfant. Cet échelonnement des évènements aux différents âges laisse plus de place pour les états intermédiaires, qui peuvent devenir plus durables.

Figure 4. Nombres moyens cumulés d'états différents vécus avant chaque âge de 15 à 35 ans. Hommes et femmes nés en 1940-44 et 1960-64

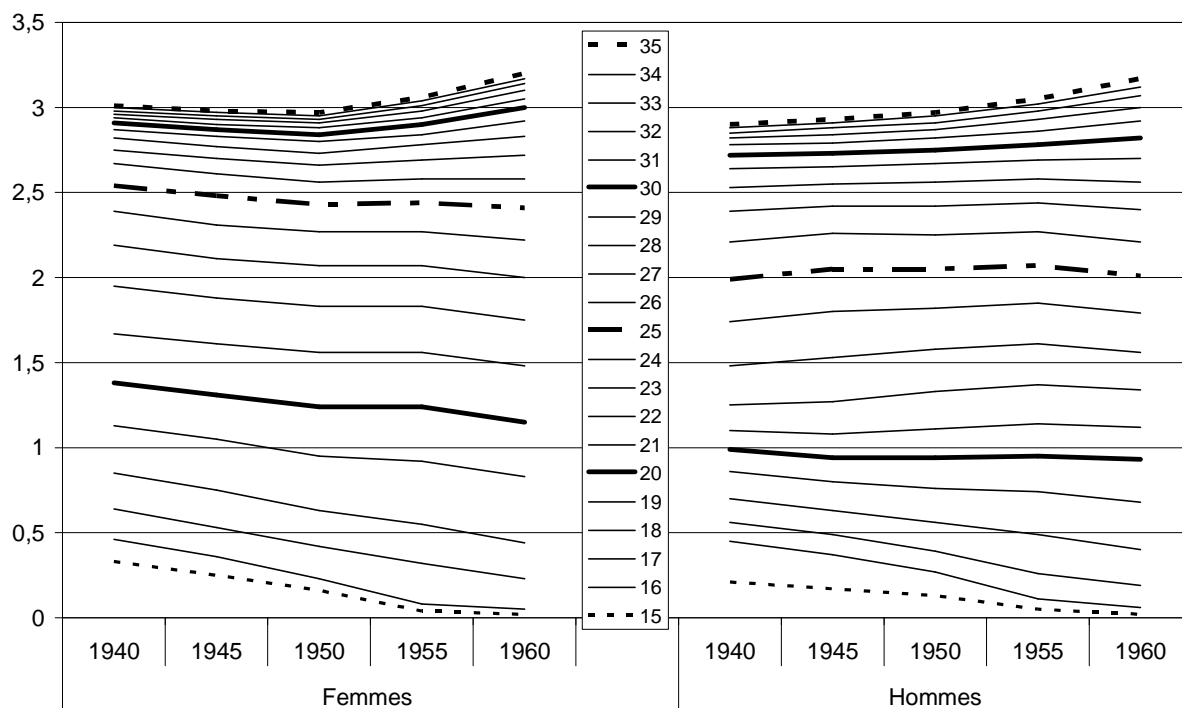
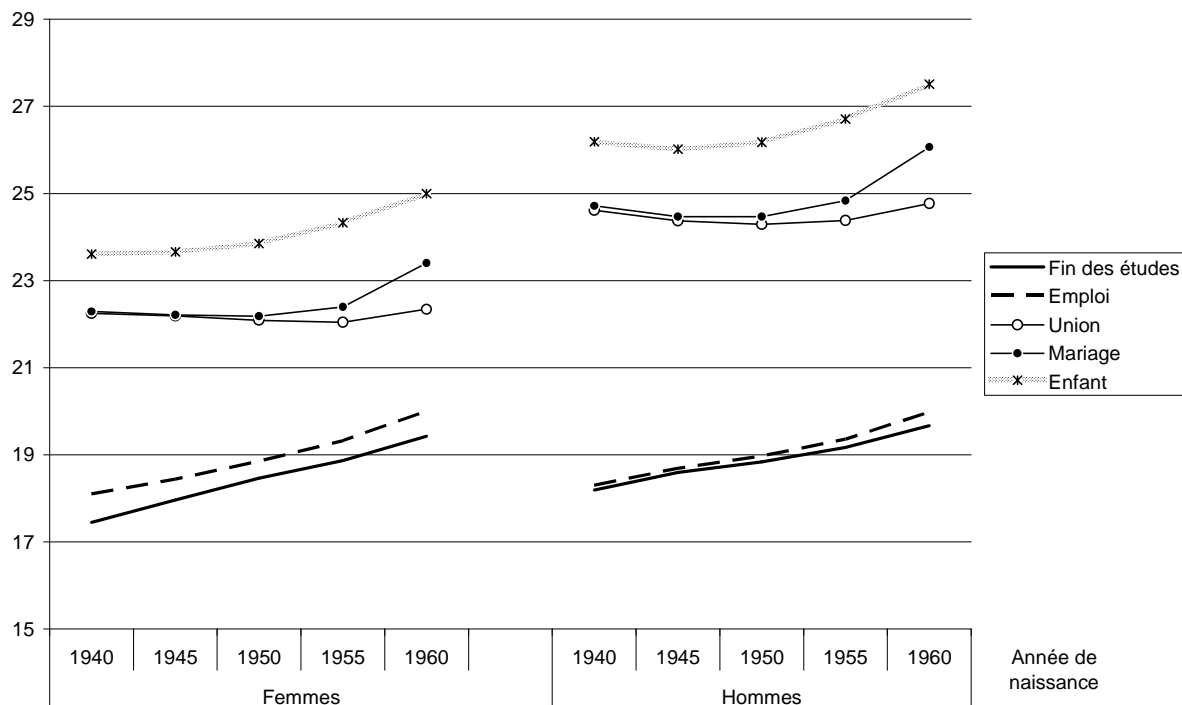


Figure 5. Âge moyen aux différents évènements. Hommes et femmes nés en 1940-44 et 1960-64



Note : les âges moyens sont calculés pour les personnes ayant vécu l'évènement entre leur quatorzième et leur trente-cinquième anniversaires.

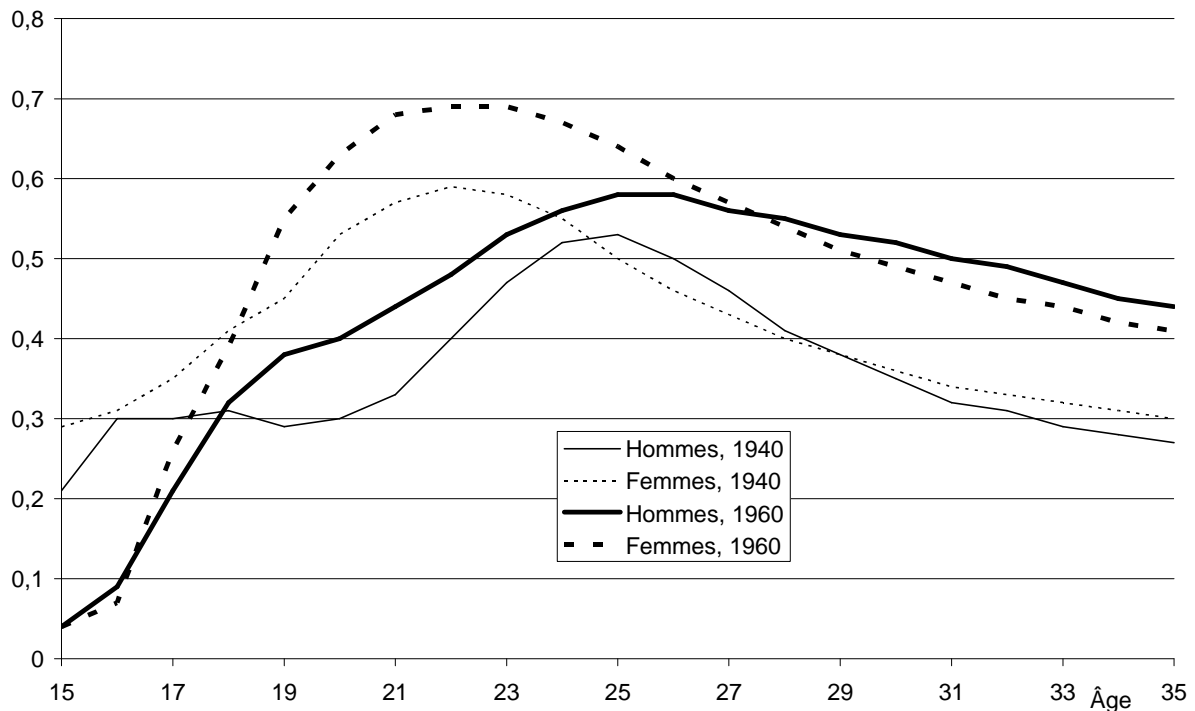
2.1.4. Entropie globale à chaque âge

Pour décrire la diversité des situations à chaque âge on utilise un critère global de dispersion des états dans lesquels se trouvent les hommes et les femmes : l'entropie, équivalent à la variance pour les variables discrètes.

$$E_t = - \frac{\sum_{j=1}^q p_{tj} \log(p_{tj})}{\log(q)}$$

Dans cette formule q représente le nombre d'états possibles (ici $q=24$) et p_{tj} est la proportion d'individus dans l'état j à l'âge t , avec par convention $0 \log(0) = 0$. On utilise ici les proportions pondérées ; c'est donc une entropie en population générale qui est estimée. L'entropie est minimale si tous les individus sont dans une même situation à l'âge t (elle vaut alors 0) et maximale si tous les p_{tj} sont égaux. Ils valent alors $1/q$, la somme vaut $\log(q)$ et l'entropie vaut 1, valeur théorique maximale pour cette définition standardisée. Pour améliorer la lisibilité de la figure 6, on n'a représenté que les générations 1940-44 et 1960-64. L'entropie varie de manière monotone d'une génération à l'autre, pour chaque sexe et groupe de génération.

Figure 6. Entropie selon l'âge, hommes et femmes nés en 1940-44 et en 1960-64



À 15 ans, l'entropie est très faible : il y a peu de diversité dans les états, parce que tous les individus présentent le même type d'étape dans leur parcours : à 15 ans la plupart des individus n'ont encore connu aucune des étapes. Dans la génération 1940 certains ont déjà fini leurs études et travaillé avant 15 ans ; pour les femmes et les hommes nés en 1960-64 très peu ont fini leurs études quasiment aucun n'a travaillé ou connu une autre des étapes observées. Pour les femmes, quelle que soit la génération l'entropie est maximum entre 22 et 23 ans. Au-delà de cet âge leurs parcours se normalisent, elles sont dans la même situation à savoir le passage de toutes les étapes ou une situation où seul le premier enfant n'a pas eu lieu. À l'opposé, l'entropie des hommes continue à augmenter après 22 ans et est maximale à 27 ans, traduisant une plus grande diversité d'états observés à cet âge là. Pour les hommes comme pour les femmes, l'entropie diminue d'une génération à l'autre avant 18 ans et augmente au-delà, surtout après 25 ans. Chez les plus jeunes les situations sont de plus en plus homogènes, à cause du retard des événements du passage vers l'âge adulte ;

après 25 ans la mise en couple n'est plus associée à un mariage et à la conception rapide d'un enfant, et les formes de vie en couple se diversifient.

Dans la génération 1940, l'entropie connaît un maximum local vers 17 ans, puis diminue jusqu'à 20 ans, avant d'augmenter fortement ensuite. Jusqu'à 25 ans, pour diminuer lentement aux âges plus élevés. C'est qu'à 20 ans les trois quarts ont fini leurs études et exercé un emploi, et très peu ont vécu en couple : à une première variabilité avant 20 ans liée aux âges différents de fin d'études et de premier emploi succède une nouvelle hétérogénéité dans les mises en couple et les étapes associées : mariage et premier enfant. Pour les femmes nées en 1940-44 comme pour les hommes nés en 1960-64, la courbe n'a qu'un maximum mais connaît une inflexion, montrant que les étapes vers l'âge adulte sont séparés en deux temps : fin des études et premier emploi d'abord, union, mariage et enfant ensuite. Pour les femmes nées en 1960-64, l'entropie augmente continûment entre 15 et 23 ans, les deux groupes d'étapes se rapprochent, notamment la fin des études et là première union (Robert-Bobée, Mazuy 2003 ; Winkler-Dworak, Toulemon 2007), ce qui explique que l'état SJ ne soit pas aussi dominant.

A titre d'exemple les tableaux 1 et 2 décrivent les situations à 22 ans, âge auquel l'entropie est maximale dans la génération 1940 pour les femmes, et à 27 ans, où elle est plus basse pour les deux sexes. Les situations les plus rares ne sont pas représentées. A 22 ans, la quasi-totalité des femmes ont déjà fini leurs études et exercé un emploi ; un tiers n'ont jamais vécu en couple ni eu d'enfant (état SJ), un cinquième sont mariées (SJUM) et autant ont, de plus, un enfant (SJUMC=A). Pour les hommes, la distribution est beaucoup plus concentrée, et l'entropie est plus faible. Deux hommes sur trois ont fini leurs études et travaillé (SJ), mais n'ont pas vécu en union ; un sur six n'a pas encore fini ses études ou jamais travaillé, un sur dix est déjà marié.

Tableau 1. Répartition des situations à 22 ans (en %). Hommes et femmes nés en 1940-44

Étapes	N	S	J	SJ	SJU	SJC	SUM	SJUC	SJUM	SJUMC	A	Autres	Ens.
Femmes	5,3	3,4	1,3	34,9	1,3	3,0	1,5	1,0	19,3	3,4	23,5	2,1	100
Hommes	11,7	3,3	2,1	66,5	1,5	1,3	0,1	0,2	7,1	0,2	4,8	1,3	100

Dans la génération 1940-44, l'entropie à 27 ans est plus importante pour les hommes que pour les femmes (Tableau 2). La moitié des hommes ont déjà connu tous les événements, un cinquième tous sauf la naissance d'un enfant, tandis qu'un quart n'ont pas vécu en couple. A cet âge deux femmes sur trois ont déjà connu tous les événements.

Tableau 2. Répartition des situations à 27 ans (en %). Hommes et femmes nés en 1940-44

Étapes	N	S	J	SJ	SJU	SJC	SUM	SJUC	SJUM	SJUMC	A	Autres	Ens.
Femmes	0,2	0,7	0,3	12,3	1,7	2,9	0,6	1,8	11,8	5,5	61,4	0,9	100
Hommes	0,6	0,8	0,4	23,9	1,9	2,1	0,3	1,7	19,4	0,2	47,1	1,6	100

On peut également décrire la variabilité des parcours par le nombre de parcours distincts pour les femmes et les hommes dans chaque groupe de génération. On souhaite en effet connaître la concentration globale des parcours : tous les individus ont-ils la même séquence répétée où existe-t-il une séquence répétée spécifique pour chaque individu ? On calcule donc le nombre de séquences répétées distinctes rapporté à l'effectif total. Plus le résultat est proche de 1 plus il existe de parcours différents, plus il est proche de 0 moins les parcours sont distincts.

Dans les générations récentes il existe une grande diversité de parcours possibles : 100 femmes nées en 1940-44 expérimentent 39 parcours différents en moyenne ; vingt générations plus tard ce sont 56 parcours différents qui sont enregistrés. Pour les hommes les parcours sont plus nombreux : 43% à 61% respectivement (Tableau 3). Les parcours rares ayant plus de chances d'apparaître dans les gros échantillons, le rapport est croissant avec de la taille de l'échantillon, toutes choses égales par ailleurs. La pluralisation des parcours est plus marquée chez les hommes que chez les femmes, bien que l'échantillon contienne moins d'hommes que de femmes. La hausse du rapport d'une génération à l'autre se fait pour chaque sexe à taille d'échantillon pratiquement constante.

On peut comparer cet indicateur à l'entropie moyenne aux âges 15 à 35 (Tableau 3, dernières lignes). Pour les femmes les parcours sont moins nombreux, mais l'inertie moyenne est plus élevée : jusqu'à

25 ans dans la génération 1940-44, jusqu'à 27 ans dans la génération 1960-64, les situations des femmes sont plus diversifiées que celles des hommes, tandis qu'aux âges plus élevés les situations finales sont plus diversifiées pour les hommes.

Tableau 3. Nombre de séquences répétées distinctes, rapporté à la taille de l'échantillon, pour les hommes et les femmes, selon la génération

	Sexe	1940-44	1945-49	1950-54	1955-59	1960-64
Effectifs	Femmes	12 309	18 163	19 851	20 019	20 801
	Hommes	8 126	12 050	13 028	13 053	13 638
Nombre de séquences distinctes	Femmes	4 816	6 623	7 833	9 284	11 730
	Hommes	3 532	5 140	6 112	6 873	8 356
Nombre de séquences distinctes / effectifs	Femmes	0,39	0,36	0,39	0,46	0,56
	Hommes	0,43	0,43	0,47	0,53	0,61
Entropie moyenne	Femmes	0,42	0,41	0,42	0,45	0,49
	Hommes	0,36	0,37	0,39	0,41	0,43

2.1.5. Diversité des séquences non répétées

Le nombre de séquences non répétées étant beaucoup plus faible que celui des séquences répétées, on peut, en faisant abstraction du temps passé dans les différents états, décrire les biographies les plus fréquentes pour les femmes et les hommes.

Tableau 4. Fréquence des séquences non répétées vécues par au moins un jeune adulte sur cent (en %), selon le sexe et la génération

		Femmes		Hommes			
		1940-44	1960-64	1940-44		1960-64	
N-SJ-SJUM-A	21,7	N-SJ-SJUM-A	11,7	N-SJ-SJUM-A	28,9	N-SJ-SJUM-A	9,8
SJ-SJUM-A	10,1	N-SJ-SJU-SJUM-A	7,5	SJ-SJUM-A	8,5	N-SJ-SJU-SJUM-A	9,4
N-S-SJ-SJUM-A	8,6	N-S-SJ-SJUM-A	5,3	N-S-SJ-SJUM-A	8,1	N-SJ	7,1
S-SJ-SJUM-A	6,6	N-SJ-SJU-SJUC	4,8	N-SJ-A	5,3	N-SJ-SJU-SJUC	5,7
N-SJ-A	3,7	N-SJ	3,4	N-SJ	3,9	N-S-SJ	5,7
SJ-A	2,7	N-S-SJ-SJU-SJUM-A	3,1	N-SJ-SJUM	3,0	N-S-SJ-SJUM-A	4,9
N-SJ	2,5	N-S-SJ-SJU-SJUC	2,8	S-SJ-SJUM-A	2,7	N-S-SJ-SJU-SJUM-A	4,8
N-S-SUM-SUMC	2,0	N-S-SJ	2,7	N-J-SJ-SJUM-A	2,6	N-SJ-SJU	4,4
N-SJ-SJUM	1,9	N-SJ-SJU	2,6	N-S-SJ	2,4	N-S-SJ-SJU-SJUC	3,4
S-SJ-A	1,8	N-SJ-SJU-SJUC-A	2,2	SJ-A	1,9	N-S-SJ-SJU	3,4
N-S-SJ	1,6	N-SJUM-A	2,1	SJ	1,6	N-SJ-SJU-SJUC-A	2,1
N-S-SJ-A	1,5	N-SJ-SJU-A	2,0	N-S-SJ-A	1,5	N-SJ-SJU-A	2,1
N-SJUM-A	1,5	N-SJU-SJUM-A	1,9	N-SJ-SJC-A	1,5	N-SJ-SJU-SJUM	1,7
SJ-SJC-A	1,3	N-S-SJ-SJU	1,7	N-SJ-SJU-SJUM-A	1,1	N-S-SJ-SJU-SJUC-A	1,6
S-SUM-SUMC	1,2	N-SJ-A	1,4	N-S-SJ-SJUM	1,1	N-J-SJ	1,4
N-J-SJ-SJUM-A	1,2	N-S-SJUM-A	1,4	N-S-SJUM-A	1,0	N-S-SJ-SJU-SJUM	1,2
N-S-SJUM-A	1,1	N-S-SJ-SJU-SJUC-A	1,1	N-SJ-SJU	0,9	N-SJ-A	1,1
N-SJ-SJC-A	1,1	N-SJ-SJU-SJUM	1,0	SJ-SJUM	0,8	N-S-SJ-SJU-A	1,1
N-S-SJ-SJUM	0,9	N-U-SJU-SJUC	1,0	N-SJUM-A	0,8	N-J-SJ-SJU-SJUM-A	1,1
S-SJ-SJC-A	0,8	N-SJU-SJUC	1,0	N-SJ-SJC	0,7	N-J-SJ-SJUM-A	1,0
SJ-SJUM	0,7	N-S-SJU-SJUM-A	0,9	N-J-SJ	0,7	N-SJ-SJUM	1,0
N-SJ-SJU-SJUM-A	0,7	N-S-SJ-SJU-A	0,9	N-SJ-SJU-SJUM	0,6	N-J-SJ-SJU-SJUC	0,8

La séquence la plus fréquente pour les femmes, N-SJ-SJUM-A, est vécue par 22% des femmes de dans la génération 1940 ; si l'on rajoute les variantes proches SJ-SJUM-A, N-S-SJ-SJUM-A et S-SJ-SJUM-A, ce sont 47% des femmes qui ont d'abord fini leur études, puis se sont mariées et ont eu un enfant (Tableau 4). En rajoutant 6% qui ont eu un enfant au même âge que leur mariage – le plus souvent une conception pré-nuptiale – le total atteint 54%. Les séquences les plus fréquentes sont

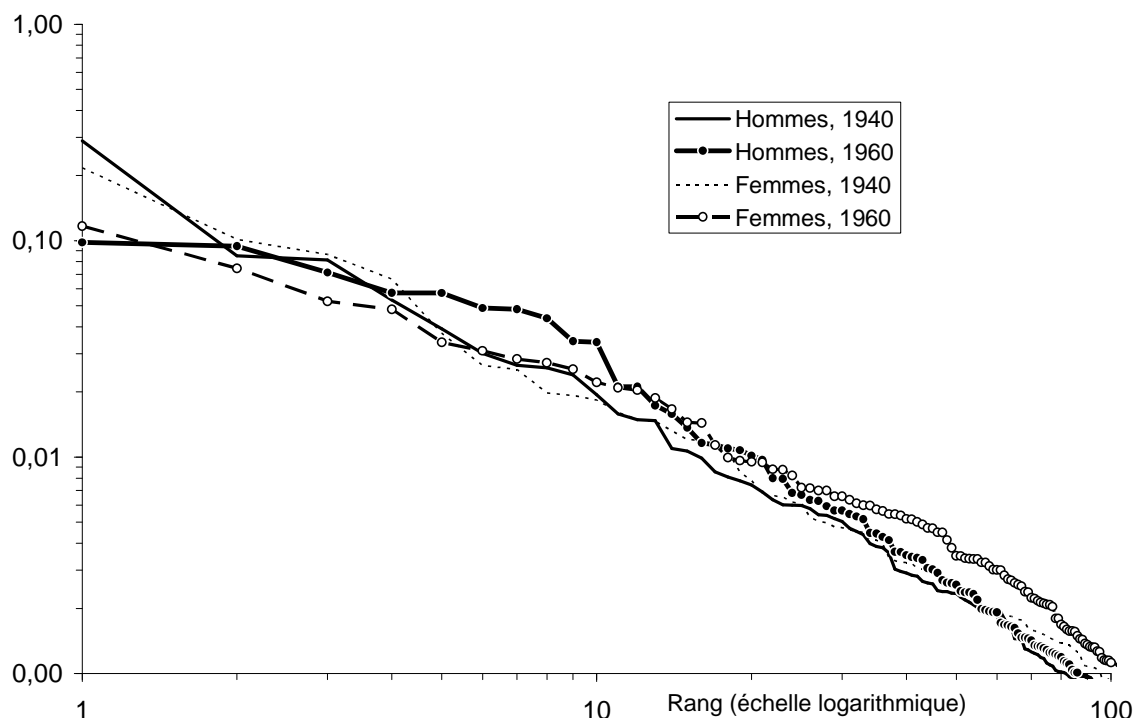
« ordonnées » (Rindfuss et al. 1987) et regroupent donc plus de la moitié des biographies. Viennent ensuite deux séquences « incomplètes » N-SJ sans union et N-S-SUM-SUMC sans emploi. Aucune des autres séquences ne regroupe plus de 2% des femmes.

Vingt générations plus tard, la séquence la plus fréquente est toujours N-SJ-SJUM-A, mais elle ne concerne plus que 12% des femmes. L'état SJU (cohabitation sans mariage après avoir travaillé) s'inclut dans les deuxième et quatrième séquences.

Chez les hommes, les séquences les plus fréquentes sont les mêmes, mais la concentration est plus importante que chez les femmes dans la génération 1940-44 : la séquence N-SJ-SJUM-A rassemble à elle seule 29% des biographies, et avec les trois mêmes variantes la proportion atteint 48%, 56% si l'on inclut également les conceptions survenues au même âge que le mariage. Les unions sont moins fréquentes pour les hommes, et les séquences incomplètes se terminant par SJ figurent en haut du tableau 4. D'une génération à l'autre, on observe la même évolution que pour les femmes, mais plus marquée : la séquence N-SJ-SJUM-A, ici encore la plus fréquente, ne concerne que moins d'un homme sur dix ; les séquences où la cohabitation remplace le mariage, soit temporairement soit jusqu'à 35 ans (N-SJ-SJU-SJUM-A et N-SJ-SJU-SJUC) occupent les deuxième et quatrième place.

Les fréquences suivent approximativement une loi rang-taille de Zipf, et les figures 7 et 8, où les séquences sont triées sur l'axe des abscisses par fréquence décroissante, montrent que dans la distribution des séquences des hommes nés en 1960-64, les séquences 4 à 10 sont « trop » fréquentes par rapport à la régularité rang-taille. C'est parce que la durée moyenne entre la fin des études et le premier emploi est faible, et selon qu'un anniversaire sépare les deux événements ou non, ils apparaissent comme simultanés ou non. On touche ici les limites de nos données, puisque les dates de ces événements ne sont pas connues au mois près, les paires de séquences différenciées uniquement par la présence ou l'absence de l'état S avant SJ, comme dans N-SJ-SJUM-A et N-S-SJ-SJUM-A, étant nombreuses parmi les séquences les plus fréquentes.

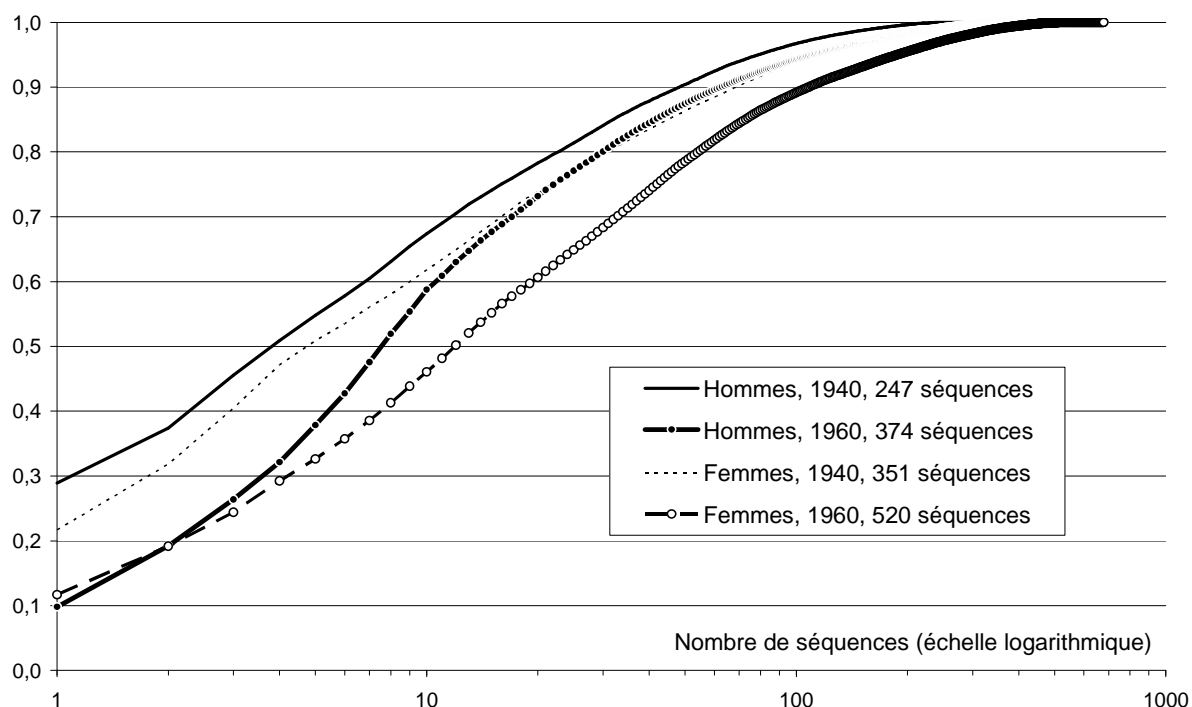
Figure 7. Fréquence des séquences non répétées classées par fréquence décroissante, hommes et femmes nés en 1940-44 et en 1960-64 (échelles logarithmiques)



Les fréquences cumulées (figure 8) montrent également la diversité croissante des biographies. On compte dans le fichier 679 séquences différentes. Pour les femmes de 1940-44, 315 séquences sont présentes dans l'échantillon, mais 4 séquences suffisent à décrire les biographies de 50% des femmes ; vingt générations plus tard il en faut 8, tandis que le nombre total de séquences différentes passe à 520. Les séquences sont plus diverses chez les femmes, puisque les nombres de séquences

nécessaires pour décrire les biographies des hommes sont de 5 et 12 dans les générations 1940-44 et 1960-64 respectivement. Les séquences sont moins nombreuses chez les hommes (247 et 374 séquences dans les générations 1940-44 et 1960-64 respectivement), mais cela vient en partie de la taille des échantillons, plus nombreux pour les femmes que pour les hommes.

Figure 8. Fréquence cumulée des séquences non répétées classées par fréquence décroissante (échelle logarithmique pour les rangs), hommes et femmes nés en 1940-44 et en 1960-64



Au total tous les indicateurs convergent pour indiquer que les séquences des femmes sont plus diverses que celles des hommes. L'ordre des événements vécus par les hommes est davantage concentré sur deux passages distincts : fin des études et premier emploi, d'abord, mise en couple, mariage et naissance du premier enfant d'autre part, tandis que pour les femmes le premier emploi peut se placer à des étapes différentes du passage à l'âge adulte.

2.2. Description des parcours individuels

On s'intéresse maintenant non plus à la diversité des parcours mais à leur complexité. Cette complexité des parcours individuels est mesurée au travers d'indicateurs synthétiques de leur variabilité, d'un âge à l'autre. On cherche ainsi à savoir si la personne expérimente de nombreux états distincts ou pas au sein de son parcours, et si les événements sont rapprochés dans le temps ou distants les uns des autres.

Pour les séquences non répétées, la complexité des parcours se résume au nombre d'états dans lesquels un individu se trouve aux différents âges, c'est-à-dire au nombre d'étapes vécues de manière non simultanée (séparées par un anniversaire), et est assez peu informative (voir figure 3 ci-dessus). Pour les séquences répétées, constituées de 21 états successifs, il existe pour mesurer la complexité des séquences individuelles deux critères : l'entropie (Shannon, 1948) et la turbulence (Elzinga, Liefbroer 2007 ; Elzinga 2006), fondés sur deux principes différents.

2.2.1. Entropie individuelle

L'entropie individuelle mesure la dispersion des différents états, à partir de la proportion π_i d'occurrences de l'état i dans la séquence x , aux âges 15 à 35. Si on note N le nombre d'états possibles (ici $N=6$ puisqu'un individu ne peut pas vivre plus de 6 états différents), on normalise l'entropie de chaque séquence à partir de la formule suivante :

$$h(x) = \frac{-\sum_{i=1}^N \pi_i \log \pi_i}{\log(N)}$$

L'entropie pour un individu est maximale s'il est passé par six états correspondant à la réalisation non simultanée des cinq événements dans son parcours de vie et s'il est resté le même temps dans chaque état. A contrario l'entropie est nulle si la séquence ne comprend qu'un état. La valeur maximale de l'entropie individuelle est 1 ; ici, avec 21 observations pour 6 états, la valeur est maximale pour 3 états présents trois fois, et 3 présents 4 fois. Examinons quelques exemples. Les entropies des séquences⁵ (S ; 2)-(SJ ; 19) et (N ; 2)-(S ; 19) sont identiques (0,18) car elles contiennent deux états et la même répartition du temps passé dans chacun des deux états. D'autre part, l'entropie de ces deux séquences est faible car presque toute la séquence est passée dans le même état. La séquence (N ; 5)-(S ; 1)-(SJ ; 4)-(SJU ; 2)-(SJUM ; 4)-(A ; 5) présente une entropie de 0,66, signifiant un parcours avec de nombreuses états et une répartition « équilibrée » du temps passé dans chaque état. Une séquence avec 6 états moins bien répartis aura une entropie inférieure. Ainsi la séquence, (N ; 4)-(S ; 1)-(SJ ; 3)-(SJU ; 1)-(SJUM ; 5)-(A ; 7) a une entropie de 0,62 bien qu'elle contienne également 6 états distincts. Dans cette séquence l'individu passe peu de temps dans les 4 premiers états, reste 5 ans entre le moment où tous les événements se sont réalisés à l'exception du premier enfant et la réalisation du premier enfant. Puis tous les événements se sont produits et il reste 7 ans dans l'état final.

Le tableau 5 présente l'évolution de l'entropie moyenne des séquences individuelles pour les hommes et les femmes des générations 1940-44 à 1960-64. La complexité des parcours augmente d'une génération à l'autre ; elle reste comparable entre hommes et femmes au sein d'une même génération, les entropies devenant identiques pour les hommes et les femmes nés en 1960-64 (Tableau 5).

Tableau 5. Entropie individuelle moyenne par sexe et génération

Génération	1940-44	1945-49	1950-54	1955-59	1960-64
Femmes	0,55	0,55	0,57	0,59	0,63
Hommes	0,57	0,57	0,59	0,62	0,64

Notons que les entropies individuelles des femmes sont légèrement inférieures à celles des hommes, alors que l'entropie de l'ensemble des femmes est plus importante que celle des hommes. Bien que les biographies des femmes ne soient en moyenne pas plus complexes que celles des hommes, entre 18 et 25 à 27 ans, les situations des femmes sont plus diverses, en raison de la plus grande variance des âges au premier emploi et de la moindre proximité entre la fin des études et le premier emploi.

2.2.2. Turbulence

Elzinga et Liefbroer (2007) ont proposé une mesure de turbulence qui prend en compte deux aspects : la variabilité du temps passé dans chaque état successif et le nombre de sous séquences non répétées distinctes. Pour une séquence x la turbulence se calcule par la formule suivante

$$T(x) = \log_2 \left(\phi(x) \frac{s_{t,\max}^2(x) + 1}{s_t^2(x) + 1} \right)$$

où $\phi(x)$ est le nombre de sous séquences non répétées distinctes de la séquence x . Par exemple la séquence répétée (S ; 2)-(SJ ; 19) correspond à la séquence non répétée S-SJ. Les sous-séquences

⁵ La séquence (S ; 2)-(SJ ; 19) est la séquence répétée S-S-SJ. On la résume en accolant à chaque état le temps passé dans cet état.

sont au nombre de 4 : par convention la séquence S-SJ elle-même et la séquence nulle, puis les séquences S et SJ. On note s_t la variance des temps passés dans chaque état, et $s_{t,max}^2$ la valeur maximale que peut prendre cette variance sachant la longueur totale de la séquence, ce maximum se calcule par :

$$s_{t,max}^2 = (n-1)(1-\bar{t})^2$$

avec n le nombre d'état distincts de la séquence et \bar{t} est la moyenne du temps passé dans chaque état, soit la longueur de la séquence divisée par le nombre d'états distincts dans la séquence. Avec $n=5$ (de N à A=SJUMC, cinq événements étant possibles dans une biographie) et \bar{t} égal à 21/6, la variance $s_{t,max}^2$ est égale à 31,25.

La turbulence est maximale pour les séquences composées de 6 états de durée égale. Nous avons choisi de normaliser la turbulence pour la comparer plus aisément à l'entropie, en la divisant par la turbulence maximale $Tmax$ ⁶.

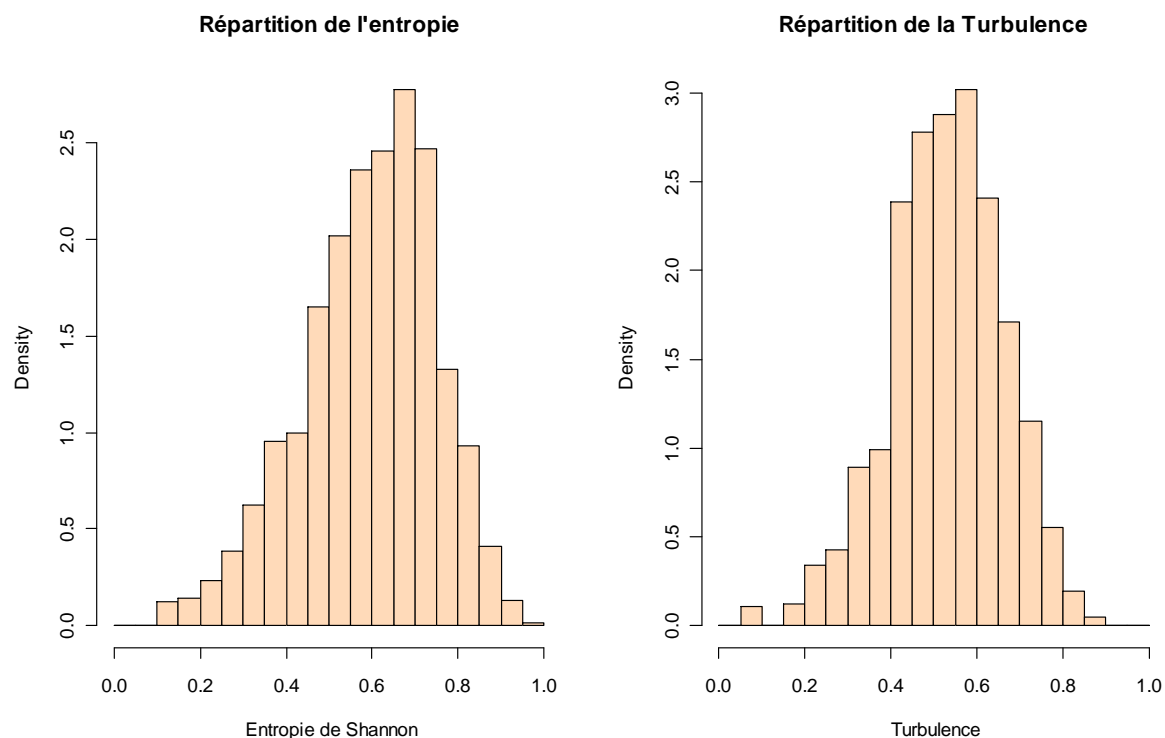
$$T(x) = \frac{\log_2 \left(\phi(x) \frac{s_{t,max}^2(x)+1}{s_t^2(x)+1} \right)}{T \max}$$

Pour des séquences qui contiennent de possibles allers-retours, la turbulence permet de tenir compte de ces allers-retours, tandis que l'entropie ne tient compte que du nombre d'occurrences de chaque état, sans distinguer si un état est présent à des âges adjacents ou séparés par un ou plusieurs autres états intermédiaires. Dans le cas des séquences telles que nous les avons définies ici, les passages sont irréversibles et cela ne se produit pas, et les deux indicateurs donnent des résultats assez proches. Les répartitions des valeurs de turbulence et d'entropie sont de la même forme, mais l'entropie est plus dissymétrique, vers les petites valeurs, tandis que la répartition des turbulences est plus symétrique (Figure 9).

La corrélation entre les valeurs de turbulence et d'entropie est très élevée : le coefficient de corrélation vaut 0,94. Pour les parcours les plus complexes, les deux indicateurs sont cependant différents. Par exemple, les séquences (N ; 5)-(SJ ; 5)-(SJUM ; 5)-(A ; 6) et (N ; 7)-(SJ ; 7)-(SJC ; 7) présentent des turbulences proches, de 0,87 et 0,84 mais des entropies de 0,77 et 0,61. En effet ces deux séquences ne présentent que peu d'états distincts (3 ou 4) donc elles ont une entropie moyenne. La turbulence, quant à elle, augmente quand les états ont des durées proches, la variance du temps passé dans chaque état est alors faible. Dans le cas de (N ; 7)-(SJ ; 7)-(SJC ; 7) cette variance est nulle et donc la turbulence est très proche de celle d'une séquence avec un état de plus.

⁶ Le maximum de la turbulence dans notre cas s'obtient pour une séquence de 6 états avec le même temps passé dans chacun des états ($s_t^2 \approx 0$). Pour 6 états, le nombre de sous-séquences est 64. Le maximum de la turbulence est donc $\log_2(64*(5*(6,25)+1))=\log_2(2064)=11,01$.

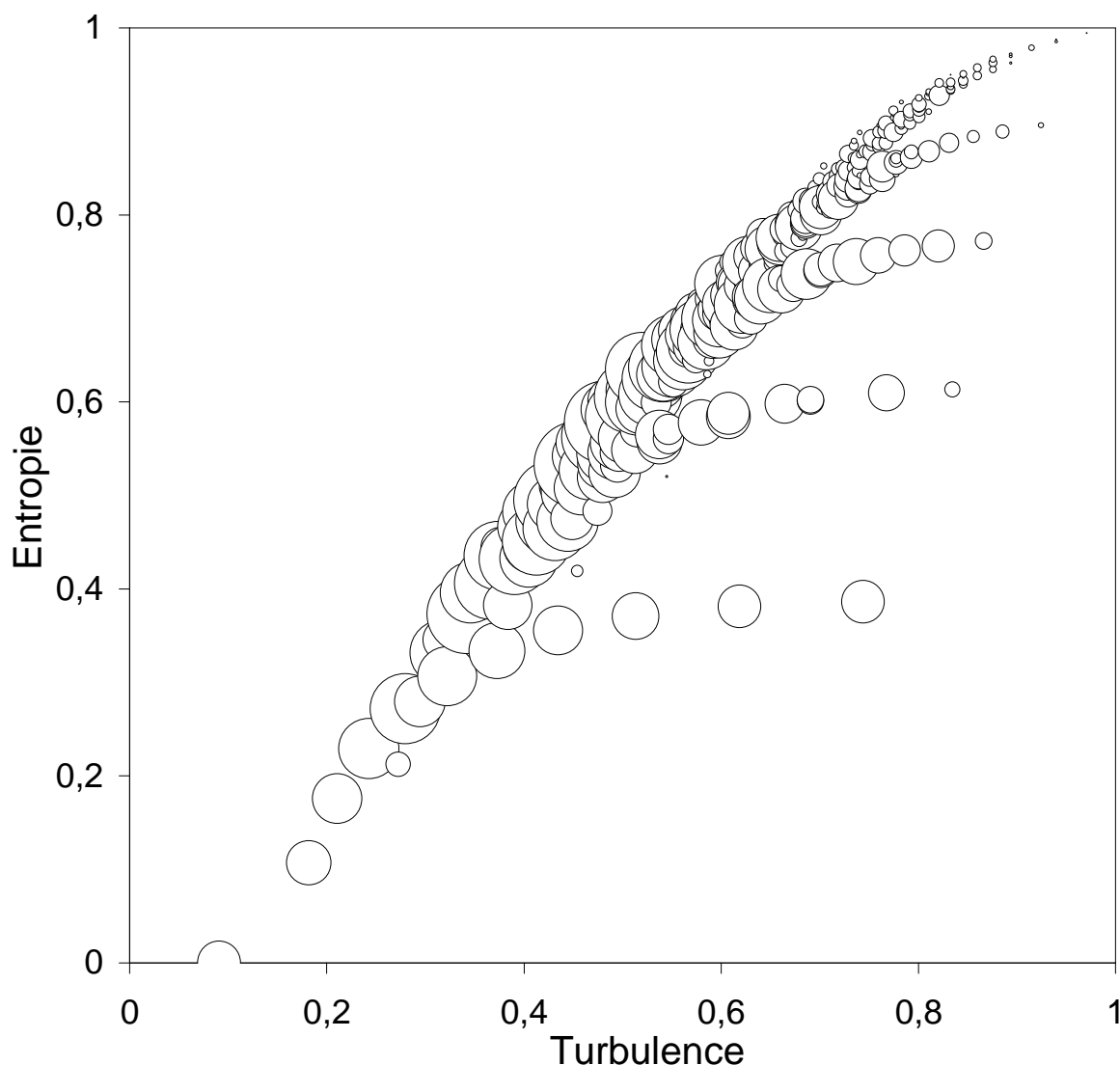
Figure 9. Répartition des valeurs d'entropies et de turbulences



L'objectif de la turbulence est de différencier des séquences pour lesquels les individus restent longtemps dans un même état et peu de temps dans les autres, séquences que l'on peut qualifier de « stables », des séquences où l'individu passe par plusieurs états avec des durées similaires ; dans ce dernier cas aucun état ne domine et les événements ne sont pas des transitions rapides vers un point stable mais sont séparés les uns des autres. Ainsi pour deux séquences dont les événements sont décalés d'un an, l'entropie sera quasiment constante mais la turbulence variera beaucoup comme dans le cas des séquences (N ; 9)-(A ; 12) et (N ; 10)-(A ; 11) (voir divers exemples au tableau annexe 1). De même, l'entropie plafonne pour un nombre d'états donné : par exemple pour 3 états les entropies sont très proches pour des nombres d'occurrence élevés, proches de l'équilibre. Ainsi des séquences (N ; 4)-(SJ ; 7)-(A ; 10), (N ; 5)-(SJ ; 7)-(A ; 9), (N ; 6)-(SJ ; 7)-(A ; 8) et (N ; 7)-(SJ ; 7)-(A ; 7) ont respectivement des entropies de 0,58, 0,60, 0,61 et 0,61. A l'inverse, leurs turbulences sont assez différentes : 0,58, 0,67, 0,77 et 0,84, puisque la variance des durées dans les différents états diminue de l'une à l'autre. La turbulence est donc proche de l'entropie si les événements sont rapprochés les uns des autres (la variance des temps passés dans les différents états est alors élevée), mais elle est plus importante si les événements sont espacés les uns des autres : dans ce cas l'entropie varie peu selon que les temps passés dans les différents états sont assez proches ou très proches, et dépend surtout du nombre d'états différents, tandis que la turbulence augmente fortement si les étapes sont espacées entre elles, donc réparties régulièrement, ce qui conduit à des durées proches passées dans les différents états. La figure 10 montre que les turbulences et les entropies des séquences de notre échantillon. En pratique les deux indicateurs sont proches, sauf pour les situations où les événements sont espacés dans les biographies.

Comme pour l'entropie, la turbulence moyenne est du même ordre de grandeur pour les hommes et pour les femmes au sein d'une même génération. Les deux indicateurs montrent une augmentation de la complexité des parcours au sein des générations les plus récentes mais peu de différence entre hommes et femmes au sein d'une même génération. En moyenne, comme pour l'entropie, la turbulence des séquences des hommes est plus élevée que celle femmes ; les deux indicateurs conduisent à la même conclusion : la différence entre les sexes diminuent très faiblement (figure 11).

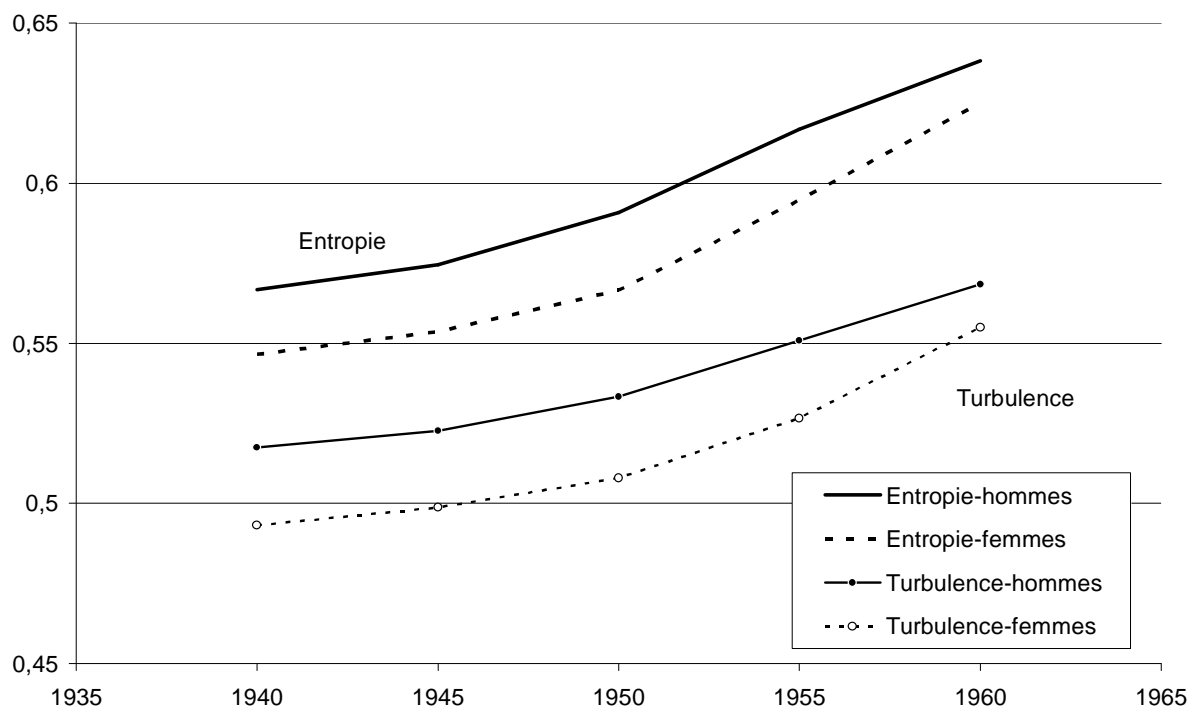
Figure 10. Entropie individuelle en fonction de la turbulence



Note : la surface de chaque cercle est proportionnelle au nombre de personnes concernées ; un cercle peut correspondre à différentes séquences.

Nicolas Robette (2008) analyse la transition vers l'âge adulte en France à partir d'une autre enquête, en utilisant des indicateurs légèrement différents (définition plus précise de la situation professionnelle, prise en compte des allers-retours dans la situation conjugale et professionnelle, prise en compte du départ de chez les parents et du nombre d'enfants) trouve une diminution de l'entropie dans les générations plus récentes pour les hommes, mais une augmentation constante de la turbulence (Robette 2008). Nos résultats semblent plus cohérents, peut-être parce qu'il calcule l'entropie à partir d'un regroupement des séquences en classes, et non à partir des séquences elles-mêmes. Par ailleurs, il trouve une turbulence plus élevée pour les femmes que pour les hommes, probablement parce qu'il définit la situation professionnelle de manière très détaillée (étudiant ; service national ; chômeur ; alternance chômage-emplois précaires ; emploi de plus de six mois ; à temps partiel ; inactif) et que les situations d'emploi des femmes sont plus variées que celles des hommes. Cette comparaison montre que les descriptions globales des niveaux et des tendances de la diversité et de la complexité des transitions vers l'âge adulte dépendent davantage des critères utilisés en amont pour définir les passages et les situations que les méthodes utilisées pour les résumer.

Figure 11. Entropie et turbulence des séquences individuelles des hommes et des femmes, générations 1940-44 à 1960-64.



3. Distance entre hommes et femmes

Dans un travail précédent sur les mêmes données, Winkler-Dworak et Toulemon (2007), utilisant des méthodes de modèles de durée, concluaient que les biographies se rapprochaient pour ce qui concerne la fin des études et le premier emploi, tandis que des différences restaient importantes pour la place de la première union dans la biographie. Des analyses de séquences partielles (comparant les évènements deux à deux) ont montré que le rapprochement des âges à la fin des études et de la mise en couple a conduit à une augmentation de la part des femmes en couple avant la fin de leurs études (Robert-Bobée, Mazuy, 2005). Ce changement explique la différenciation plus importante des biographiques féminines, en plus des facteurs communs aux deux sexes, concernant la diffusion de la cohabitation hors mariage.

Pour résumer le rapprochement ou la différenciation des parcours entre les hommes et les femmes, ce qui était notre question de départ, nous avons essayé un grand nombre d'indicateurs synthétiques fondés sur les séquences non répétées et répétées. La mixité des séquences offre une première mesure de la proximité entre hommes et femmes : dire que les séquences sont franchies par le même nombre d'hommes et de femmes revient à dire que femmes et hommes franchissent les mêmes étapes au même moment. Les distances entre hommes et entre femmes d'une part, entre hommes et femmes d'autre part, fournissent un deuxième indicateur : si les séquences de deux hommes ou de deux femmes ne sont en moyenne pas plus différentes que les séquences d'un homme et d'une femme, c'est que la différenciation entre hommes et femmes est peu pertinente, donc que les parcours des hommes et des femmes sont similaires.

3.1. Mesurer la mixité des séquences

À partir des séquences non répétées, l'indicateur le plus simple de mixité globale consiste à faire la somme, pour chaque séquence et chaque génération, de la fréquence minimale de la séquence parmi les femmes et parmi les hommes : si les répartitions des femmes et des hommes entre les séquences se rapprochent, cette somme augmente. D'une génération à l'autre, la proportion diminue très légèrement, de 77,3% à 76,4%, entre les générations 1940-44 et 1960-64.

Il est donc nécessaire de décrire l'évolution de la mixité des différentes séquences. Cela peut se faire en utilisant un indicateur de mixité pour chaque séquence, $Mix(i)$, défini comme le rapport des chances (*odds ratio*) de la fréquence de la séquence chez les femmes et chez les hommes, si la séquence est plus fréquente chez les hommes, et de son inverse sinon.

$$OR(h,i) = \frac{p_h(i)}{1 - p_h(i)}$$

$$OR(f,i) = \frac{p_f(i)}{1 - p_f(i)}$$

$$Mix(i) = \min\left(\frac{OR(h,i)}{OR(f,i)}; \frac{OR(f,i)}{OR(h,i)}\right)$$

Par construction, $Mix(i)$ est compris entre 0 et 1 : $Mix(i)$ vaut 0 si la séquence i est présente pour un seul sexe, et vaut 1 si sa fréquence chez les femmes et les hommes sont identiques. L'utilisation des *odds ratios* permet de comparer des séquences de fréquence très différentes.

L'évolution de la mixité se mesure alors facilement par le rapport des mixités pour deux générations ⁷.

$$\Delta Mix(i, g_1, g_2) = \frac{Mix(i, g_2)}{Mix(i, g_1)}$$

On peut calculer ces indicateurs pour les 98 séquences dont la fréquence dépasse 0,3% pour au moins une génération, pour les hommes ou pour les femmes. Ces séquences regroupent plus de 90% des individus de la génération 1960-64 (voir figure 8) ⁸. Pour 46 séquences, regroupant 49% de la génération 1960-64, la mixité augmente ; pour 52 séquences et 41% des individus, elle diminue. Selon l'indicateur choisi, on conclut donc soit à une hausse soit à une baisse de la mixité, mais dans les deux cas les évolutions sont très faibles.

3.2. Des évolutions très diverses selon les séquences

L'examen des séquences dont la mixité augmente des générations 1960-64 aux générations 1940-44 peut se faire en comparant les séquences selon plusieurs critères. L'évolution de la fréquence d'une séquence est représentée par l'*odds ratio* de la fréquence de la séquence tous sexes confondus ; l'évolution de la mixité est représentée par le rapport Mix et son évolution. Ces analyses sont résumées sur la Figure 12. Des figures complémentaires sont présentées en annexe 2.

Les séquences commençant par N-SJ-SJU (fin des études et premier emploi, puis mise en couple sans mariage, suivi d'un mariage, d'une naissance ou des deux, deviennent à la fois plus fréquentes et plus mixtes (en haut à droite sur la figure 12) ; ces séquences étaient plutôt masculines dans la génération 1940-44, et leur fréquence a augmenté surtout chez les femmes (Figure annexe 1) et la plus fréquente d'entre elles est N-SJ-SJU-SJUM-A. Le quadrant supérieur gauche de la figure 12 regroupe les séquences qui deviennent plus mixtes, mais moins fréquentes : la sous-séquence SJ-SJUM décrivant un mariage direct (sans cohabitation préalable) les caractérise. On retrouve ici la séquence la plus fréquente, N-SJ-SJUM-A et ses variantes, décrites en 2.1.5.

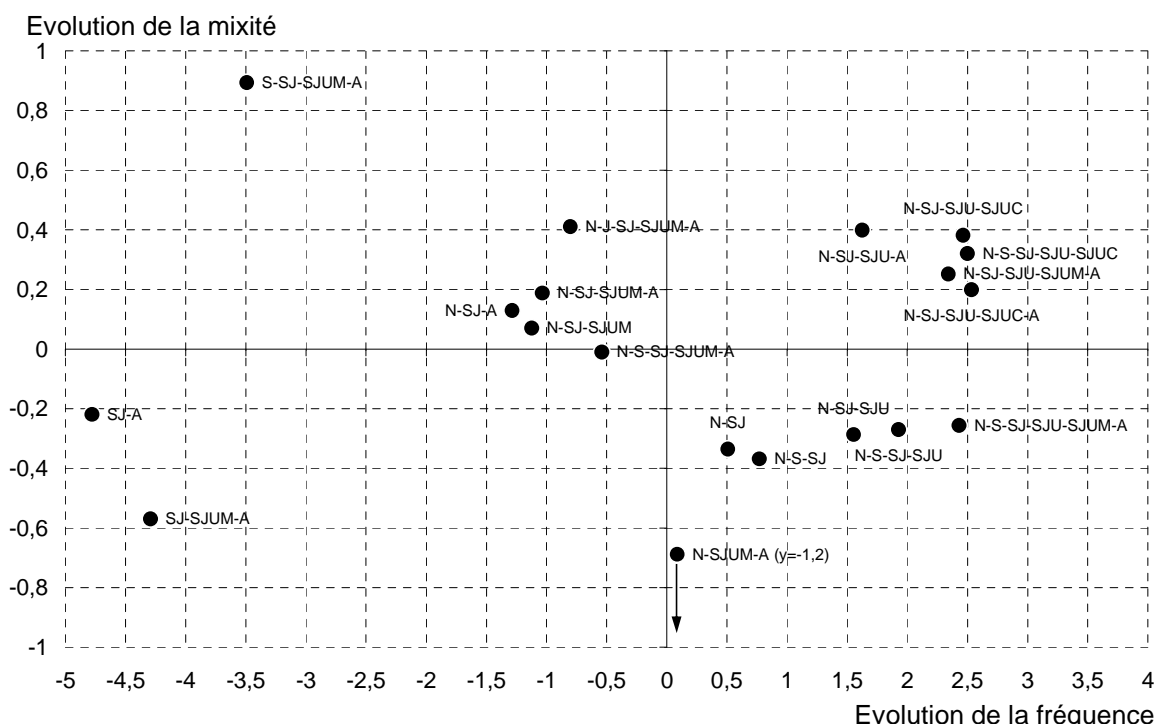
En bas à droite de la figure 12, les séquences commençant par N-SJ ou N-S-SJ et se terminant soit par l'absence d'union, soit par une union sans mariage ni naissance, deviennent moins mixtes et plus fréquentes : caractéristiques des biographies des hommes, elles deviennent plus fréquentes chez eux, sans se développer chez les femmes.

⁷ L'évolution de la mixité peut s'accompagner éventuellement d'un changement du sexe le plus représenté parmi les individus ayant connu cette séquence.

⁸ Le choix d'autres seuils donne des résultats très similaires.

Enfin, les séquences SJ-SJUM-A et SJ-A sont isolées en bas à gauche de la figure 10 : très fréquentes dans la génération 1940-44, elles deviennent beaucoup moins fréquentes et moins mixtes. Ces deux séquences étaient plutôt féminines dans la génération 1940-44, elles deviennent plutôt masculines dans les générations récentes, par désaffection plus marquée des femmes pour ces séquences très fortement clivées : fin des études et premier emploi dans un premier temps avant 15 ans, union, mariage et premier enfant ensuite, chaque transition regroupant des événements dans un court laps de temps. On peut rapprocher de ce groupe la séquence N-SJUM-A dans laquelle tous les événements (à part la naissance du premier enfant) ont lieu en même temps, la fin des études et le premier emploi étant moins précoce. Cette séquence est rare et devient de plus en plus féminine.

Figure 12. Évolution de la mixité des séquences en fonction de l'évolution de la fréquence des séquences. Logarithme de l'odds ratio de la fréquence en 1960-64 et en 1940-44 pour les deux sexes en abscisses. Logarithme du rapport de la mixité dans la génération 1960-64 à la mixité en 1940-44 en ordonnées.



Note : Le point N-SJUM-A serait hors du graphique à cette échelle, l'ordonnée a été remontée de -1.2 à -0.67.

Une classification ascendante hiérarchique⁹ opérée sur les séquences non répétées (non représentée ici) conduit aux mêmes résultats, les groupes de séquences correspondant aux quadrants décrits à la figure 12. Les séquences suivent donc des évolutions très contrastées pour chaque sexe, selon les sous-séquences qu'elles contiennent.

3.3. Les séquences des hommes et des femmes ne deviennent pas plus similaires

Au terme de cette analyse des séquences non répétées, la conclusion s'impose : la diversification des séquences s'accompagne d'une hausse de l'entropie et de la turbulence, un peu plus élevée pour les femmes que pour les hommes, mais les séquences ne deviennent pas plus mixtes en moyenne.

Les changements importants dans le calendrier de la fin des études et du premier emploi, la diffusion de la vie en couple sans mariage et le retard des premières naissances ont entraîné des changements

⁹ La classification a été réalisée à partir de la matrice de distance obtenue par optimal matching avec un coût de un pour l'insertion-suppression et de deux pour la substitution. La classification a été faite sur les séquences non répétées distinctes chaque séquence distincte ne comptant que pour un quelque soit le nombre d'individus l'ayant réalisée.

importants dans les calendriers des différents événements. Les séquences répétées permettent de compléter l'analyse des séquences non répétées pour tenir compte de la durée passée dans chaque état. Pour cela nous avons calculé, pour toutes les paires d'individus (a,b) retenus, la distance entre a et b $d(a,b)$ à partir d'un optimal matching¹⁰, puis la contribution de la distance entre a et b à l'inertie totale, à l'aide des poids de a et b , poids p_a et p_b normalisés pour que, dans chaque génération, le poids total soit de un :

$$I(a,b) = p_a^2 p_b^2 d^2(a,b)$$

Puis l'inertie globale dans chaque génération est décomposée en trois composantes ; l'inertie des hommes $I(H)$, l'inertie des femmes $I(F)$ et l'inertie des couples hommes-femmes $I(m)$:

$$I(H) = \frac{\sum_{(h_1, h_2) \in H^2} p_{h_1}^2 p_{h_2}^2 d^2(h_1, h_2)}{\left(\sum_{h \in H} p_h \right)^2}$$

$$I(F) = \frac{\sum_{(f_1, f_2) \in F^2} p_{f_1}^2 p_{f_2}^2 d^2(f_1, f_2)}{\left(\sum_{f \in F} p_f \right)^2}$$

$$I(m) = \frac{\sum_{h \in H} \sum_{f \in F} p_h^2 p_f^2 d^2(h, f)}{\left(\sum_{f \in F} p_f \right) \left(\sum_{h \in H} p_h \right)}$$

$$I = I(H) + I(F) + 2I(m)$$

Finalement, la part de l'inertie totale imputable aux différences hommes femmes peut se mesurer par le rapport de l'inertie des couples homme-femme à l'inertie totale

$$R^2 = \frac{2I(m)}{I}$$

Si les effectifs de femmes et d'hommes sont identiques et si les hommes et les femmes sont semblables en moyenne, la variable « sexe » n'apporte aucune information et R^2 vaut 0,50. Si les hommes et les femmes représentent des groupes homogènes et différents l'un de l'autre, R^2 vaut 1.

Tableau 6. Variabilité des séquences entre hommes, entre femmes et entre hommes et femmes

Génération	Variabilité hommes $I(H)$	Variabilité femmes $I(F)$	Variabilité mixte $I(m)$	Variabilité totale I	R^2
1940-44	247	309	294	1144	0,5144
1960-64	305	356	350	1361	0,5142

On retrouve avec cet indicateur l'augmentation, d'une génération à l'autre, de la diversité des séquences déjà observées à partir de l'inertie, plus marquée pour les femmes, dont l'inertie totale est supérieure à celle des hommes. Le R^2 reste remarquablement constant (tableau 6). Mais, surtout, la valeur du R^2 est proche de 50%, preuve que le sexe n'explique qu'une très faible partie de la diversité des séquences. Peut-être d'autres coûts de substitution, plus faibles, rapprochant des séquences

¹⁰ Optimal matching avec un coût de un pour l'insertion-suppression et de deux pour la substitution

répétées correspondant à la même séquence non répétée, mais avec des durées différentes dans les différents états, rendraient mieux compte des différences entre femmes et hommes. Cependant le choix des coûts de substitution reste largement arbitraire.

Conclusion

On peut décrire le passage à l'âge adulte comme le franchissement d'un certain nombre d'étapes : fin des études, premier emploi, mise en couple, mariage, naissance du premier enfant. À partir de la date de ces événements on peut définir, à chaque âge, l'état dans lequel est chaque individu à partir des événements qu'il a déjà vécu. La suite des états aux âges 15 à 35 définit une séquence, que l'on peut résumer par la suite des états avec leur durée (séquences répétées) ou sans préciser leur durée (séquence non répétées).

En utilisant l'enquête Étude de l'histoire familiale de 1999, nous avons ainsi décrit le passage à l'âge adulte en France dans les générations du début des années 1940 et des années 1960.

Le résultat principal est la diversification accrue des parcours et l'augmentation de la turbulence des biographies : les histoires individuelles sont faites de plus d'états différents, bien que moins d'événements soient vécus avant 35 ans, à cause de la désynchronisation des passages, dans les générations nées au début des années 1960, par rapport aux générations nées 20 ans plus tôt. On observe moins d'événements simultanés, et de nouveaux états intermédiaires enrichissent les séquences, comme la cohabitation sans mariage ni enfant. La diversité s'accroît également en termes d'ordre dans lequel les événements sont vécus, ce qui augmente les états différents expérimentés par les jeunes, d'une génération à l'autre.

Tous les indicateurs globaux (entropie des états, entropie individuelle, turbulence des séquences) convergent pour confirmer ce constat. Par contre, ces indicateurs globaux ne fournissent pas une réponse claire à la question sur l'évolution des différences entre hommes et femmes. Les parcours féminins sont plus diversifiés mais moins complexes que ceux des hommes. Tout se passe comme si d'une part le champ des parcours possibles était plus faible pour les femmes mais comme si d'autre part les femmes modulaient davantage leur parcours dans ce champ des possibles. Pour les hommes comme pour les femmes, tous les indicateurs indiquent donc une diversification et une complexification des parcours, sans que l'on puisse dire que les parcours masculins et féminins se rapprochent au fil du temps. D'une génération à l'autre, la « mixité » des séquences n'augmente pas. Peut-être la convergence a-t-elle lieu dans les générations plus jeunes, non observées ici, mais tout laisse à penser que les évolutions récentes n'ont pas conduit à une telle convergence : le rapprochement des âges à la fin des études et à la mise en couple a entraîné une augmentation des unions avant le premier emploi pour les femmes, mais pas pour les hommes, comme si ces derniers restaient réticents face à ce « désordre » dans les transitions, le premier emploi étant pour eux une étape préalable à la mise en couple (Robert-Bobée et Mazuy 2005, Winkler-Dworak et Toulemon 2007).

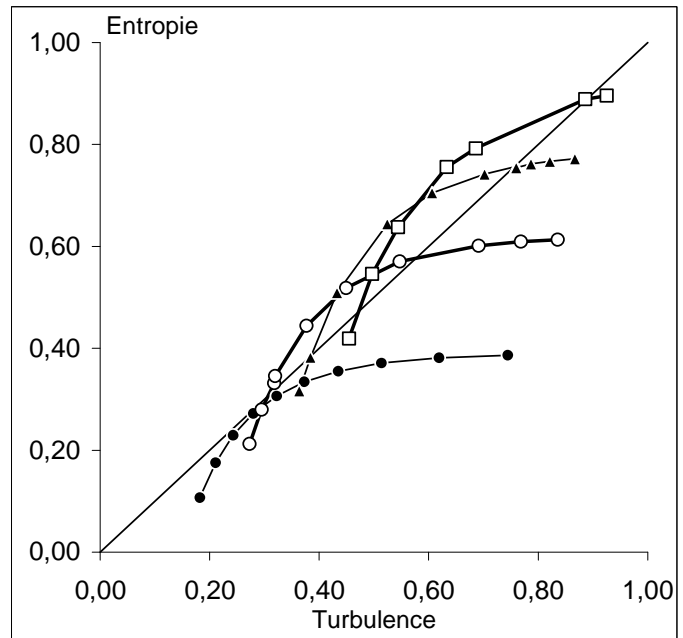
Ces conclusions sont très robustes aux méthodes choisies, mais fortement dépendantes du choix des indicateurs du passage vers l'âge adulte. Robette (2008), dans un travail utilisant le départ de chez les parents et analysant la situation d'emploi à partir d'une nomenclature fine, arrive à des conclusions assez différentes des nôtres (parcours plus turbulents pour les femmes, entropie non croissante pour les dernières générations). L'analyse des séquences elles-mêmes, voire des sous-séquences et de l'ordre temporel des paires d'événements (Toulemon et al. 2007), permet d'affiner l'analyse sans se limiter à des résultats globaux.

Annexes

Tableau annexe 1. Comparaison de l'entropie et de la turbulence pour quelques séquences

Le tableau ci-dessous présente les valeurs de la turbulence (en abscisses) et de l'entropie (en ordonnées) pour quelques séquences répétées. Pour faciliter la lecture, les états ont été nommés A, B, C, D et E.

Etats	Sequence	Turbulence	Entropie
2	(A,1)-(B,20)	0,18	0,11
2	(A,2)-(B,19)	0,21	0,18
2	(A,3)-(B,18)	0,24	0,23
2	(A,4)-(B,17)	0,28	0,27
2	(A,5)-(B,16)	0,32	0,31
2	(A,6)-(B,15)	0,37	0,33
2	(A,7)-(B,14)	0,43	0,36
2	(A,8)-(B,13)	0,51	0,37
2	(A,9)-(B,12)	0,62	0,38
2	(A,10)-(B,11)	0,74	0,39
3	(A,1)-(B,1)-(C,19)	0,27	0,21
3	(A,1)-(B,2)-(C,18)	0,29	0,28
3	(A,1)-(B,3)-(C,17)	0,32	0,33
3	(A,2)-(B,2)-(C,17)	0,32	0,35
3	(A,3)-(B,3)-(C,15)	0,38	0,44
3	(A,4)-(B,4)-(C,13)	0,45	0,52
3	(A,5)-(B,5)-(C,11)	0,55	0,57
3	(A,5)-(B,8)-(C,8)	0,69	0,60
3	(A,6)-(B,7)-(C,8)	0,77	0,61
3	(A,7)-(B,7)-(C,7)	0,84	0,61
4	(A,1)-(B,1)-(C,1)-(D,18)	0,36	0,32
4	(A,1)-(B,1)-(C,2)-(D,17)	0,38	0,38
4	(A,2)-(B,2)-(C,2)-(D,15)	0,43	0,51
4	(A,3)-(B,3)-(C,3)-(D,12)	0,52	0,64
4	(A,3)-(B,4)-(C,4)-(D,10)	0,61	0,70
4	(A,3)-(B,5)-(C,5)-(D,8)	0,70	0,74
4	(A,4)-(B,4)-(C,6)-(D,7)	0,76	0,76
4	(A,3)-(B,6)-(C,6)-(D,6)	0,76	0,75
4	(A,4)-(B,5)-(C,5)-(D,7)	0,79	0,76
4	(A,4)-(B,5)-(C,6)-(D,6)	0,82	0,77
4	(A,5)-(B,5)-(C,5)-(D,6)	0,87	0,77
5	(A,1)-(B,1)-(C,1)-(D,1)-(E,17)	0,45	0,42
5	(A,1)-(B,1)-(C,2)-(D,2)-(E,15)	0,50	0,55
5	(A,1)-(B,1)-(C,3)-(D,3)-(E,13)	0,54	0,64
5	(A,1)-(B,2)-(C,4)-(D,4)-(E,10)	0,63	0,76
5	(A,1)-(B,2)-(C,5)-(D,5)-(E,8)	0,69	0,79
5	(A,3)-(B,4)-(C,4)-(D,5)-(E,5)	0,89	0,89
5	(A,4)-(B,4)-(C,4)-(D,4)-(E,5)	0,93	0,90

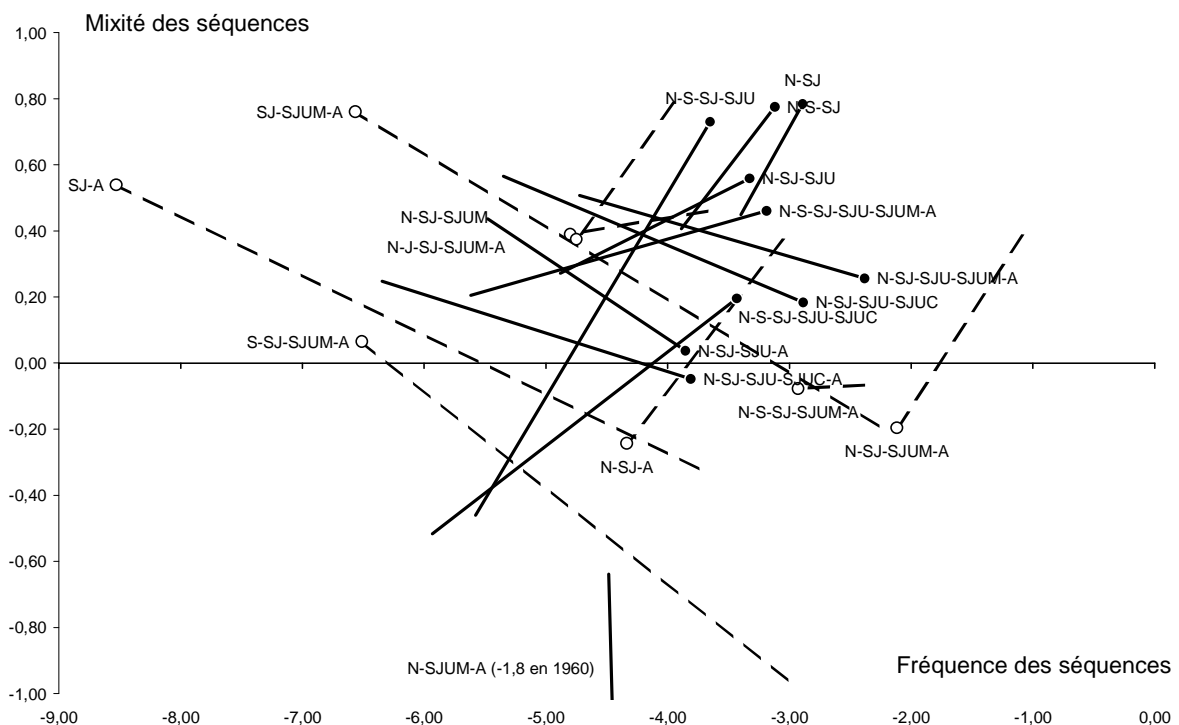


Pour un nombre fixé d'états (ici de 2 à 5), l'entropie et la turbulence sont proches quand quelques états sont peu durables, l'entropie étant plus élevée que la turbulence quand un état domine fortement les autres. Puis l'entropie augmente peu dès que chaque état est présent sur 20% des âges, tandis que la turbulence continue d'augmenter au fur et à mesure que la variance des durées dans chaque état diminue, c'est-à-dire que les états sont également durables, ou encore que les événements qui séparent les états sont régulièrement espacés entre 15 et 35 ans. L'entropie est donc plus sensible au nombre d'états différents, tandis que la turbulence accorde plus d'importance à l'espacement des événements, qui se mesure par la (faible) variance des durées dans chaque état.

Annexe 2 : mixité et fréquence des séquences non répétées

Sur la figure annexe 1, chaque séquence est représentée par un trait, qui représente l'évolution de la génération 1940-44 à la génération 1960-64 (repérée par un point). En abscisse, la fréquence de la séquence est indiquée par le log-odds ratio de la fréquence moyenne hommes-femmes ; en ordonnée la masculinité est indiquée par le log-odds ratio de la fréquence chez les hommes, relativement aux femmes. Les séquences qui se « dirigent » vers la gauche deviennent moins fréquentes avec le temps, tous sexes confondus ; celles qui se dirigent vers le haut deviennent plus masculines ; les séquences qui se rapprochent de la barre horizontale (ordonnée nulle représentant la mixité de la séquence) deviennent plus mixtes, celles qui s'en éloignent deviennent plus masculines (vers le haut) ou féminines (vers le bas).

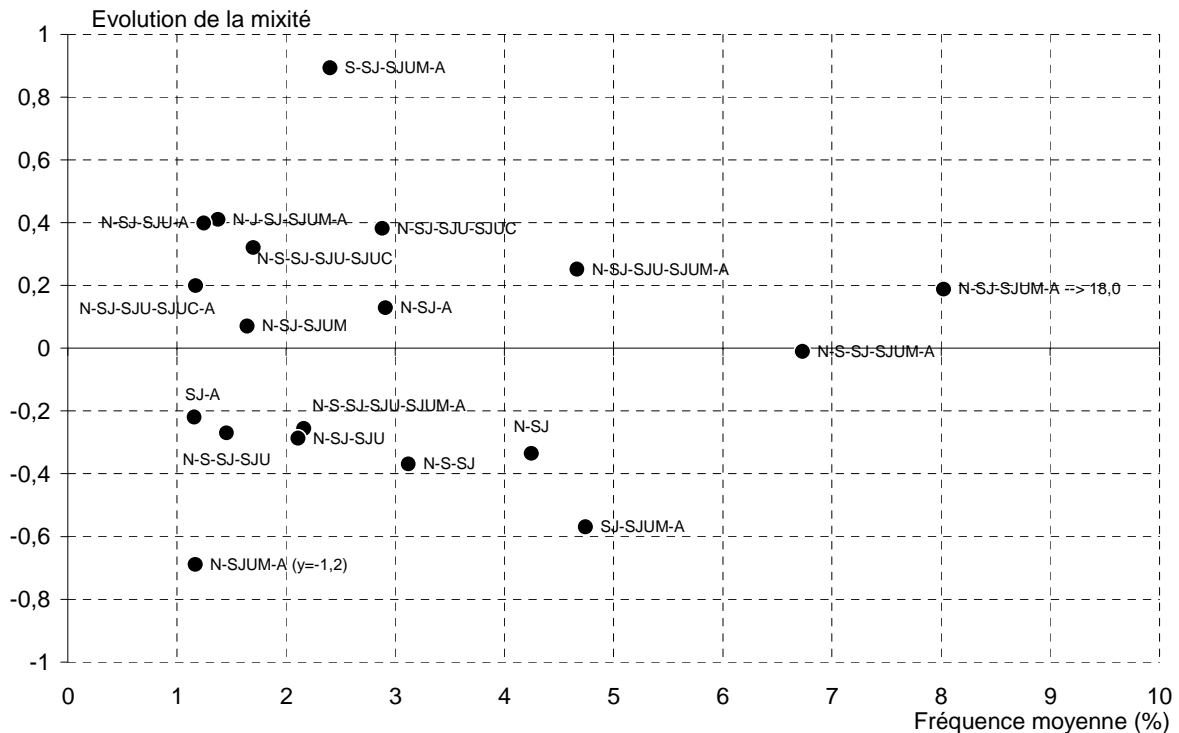
Figure annexe 1. Fréquence des séquences non répétées (log-odds ratio de la fréquence moyenne hommes-femmes) en abscisses et masculinité des séquences (log-odds ratio de la fréquence chez les hommes, relativement aux femmes) en ordonnées. Les traits représentent la position de chaque séquence dans les générations 1940-44 et 1960-64 (identifiées par un point)



Note : Le point N-SJUM-A est hors du graphique à cette échelle, l'ordonnée en 1960 vaut -1,8.

La figure annexe 2 présente l'évolution de la mixité en ordonnées (même information que dans la figure Annexe 1), en fonction de la fréquence moyenne de la séquence, toutes générations et tous sexes confondus. Elle permet de relativiser les évolutions en isolant les séquences les plus fréquentes (à droite du graphique). La séquence N-SJ-SJUM-A concerne 18% des personnes, le point correspondant a été déplacé pour trouver sa place sur le graphique

Figure annexe 2. Fréquence des séquences non répétées (fréquence moyenne hommes-femmes, en %) en abscisses et masculinité des séquences (log-odds ratio de la fréquence chez les hommes, relativement aux femmes) en ordonnées.



Note : Le point N-SJUM-A serait hors du graphique à cette échelle, l'abscisse a été ramenée de 18% à 8%.

Bibliographie

- [1] Abbott A., Forrest J., 1986, "Optimal Matching Methods for Historical Sequences", *Journal of Interdisciplinary History*, 26, p. 471-494
- [2] d'Aubigny, G., 1993. « Analyse des proximités et programmes de codage multidimensionnel ». *Modulad*, 12, p. 1-32.
- [3] Billari F., Piccarreta R., 2005, "Analysing Demographic Life Courses Through Sequence Analysis", *Mathematical Population Studies*, vol. 12, n°2, p. 81-106.
- [4] Billari F., 2001, "The analysis of early life courses: complex descriptions of the transition to adulthood", *Journal of Population Research*, vol. 18, n°2: pp.119-142.
<http://www.jpr.org.au/upload/JPR18-2pp118-142.pdf>
- [5] Cassan F., Héran F., Toulemon L., 2000, « Étude de l'histoire familiale : l'édition 1999 de l'enquête Famille », *Courrier des statistiques*, n°93, p. 19-23.
- [6] Courgeau D., Guerin-Pace F., 1998, Le suivi des itinéraires professionnels des couples par les méthodes de la statistique textuelle. Lecture des parcours professionnels des couples. – dans : JADT

1998, 4èmes journées internationales d'analyse des données Textuelles, Université de Nice – Sophia Antipolis, p. 221-232. <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt1998/courgeau.htm>

[7] Elzinga, C., 2006. "Sequence analysis: Metric representation of categorical time series". *Sociological Methods & Research*, under revision.

[8] Gabadinho, A., G. Ritschard, M. Studer and N. S. Müller Mining sequence data in R with the TraMineR package: A user's guide University of Geneva, 2008. (<http://mephisto.unige.ch/traminer>)

[9] Galland O., 2000, « Entrer dans la vie adulte : des étapes toujours plus tardives, mais resserrées », *Economie et statistique*, n°337-338, p. 13-36.

[10] Galland O., 2007, *Sociologie de la jeunesse. L'entrée dans la vie*, Paris, Armand Colin, collection U, série «Sociologie».

[11] Kruskal, Joseph B., 1983, « An overview of sequence comparison », in David Sankoff and Joseph B. Kruskal. *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. Reading, MA: Addison-Wesley, p. 1-44.

[12] Lefèvre C., Filhon A. (dir.), 2005, *Histoires de familles, histoires familiales. Les résultats de l'enquête Famille de 1999*, Les Cahiers de l'INED, n°156. Paris : Ined, XXXV III+642 pages.

[13] Lesnard L., de Saint Pol T., 2006,. « Introduction aux méthodes d'appariement optimal (optimal matching analysis) ». *Bulletin de méthodologie sociologique*, 90. Mis en ligne le 01 avril 2009. <http://bms.revues.org/index638.html>.

[14] Rindfuss R., Swicegood G., Rosenfeld R., 1987, "Disorder in the Life Course: How Common and Does It Matter". *American Sociological Review*, vol. 52, n°6, p. 785-801.

[15] Robert-Bobée I., Mazuy M., 2003. Calendriers de constitution des familles et âge de fin des études. Chapitre 8 in Lefèvre C., Filhon A. (eds.), 2005, *Histoires de familles, histoires familiales. Les résultats de l'enquête Famille de 1999*, Les Cahiers de l'INED, N°156. PARIS: INED, p. 175-200.

[16] Robette N., 2008, The diversity of pathways to adulthood: a life course typologies approach, communication au European Population Conference, EAPS, Barcelone, 11 Juillet 2008. <http://epc2008.princeton.edu/download.aspx?submissionId=80073>

[17] Saporta G., 2006. *Probabilités, analyse des données et statistique*. 2e édition révisée et augmentée. Paris.

[18] Shannon, C., 1948, *A mathematical theory of communication* Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, <http://pespmc1.vub.ac.be/books/Shannon-TheoryComm.pdf>

[19] Schoenberg J., 1937. On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Annals of mathematics*, vol. 38, n°4, p. 787–793.

[20] Toulemon L., 1996, « La cohabitation hors mariage s'installe dans la durée », *Population*, vol. 51, n°3, p. 675-716.

[21] Toulemon L., Winkler-Dvorak M., Morand E., 2008, « Gender differences in the transition to adulthood in France: Is there convergence over the recent period ? Hazard models and sequence analysis ». Seminar on early adulthood in time and space: multidimensional and multilevel perspectives IUSSP panel Ined, Paris

[22] 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*, Vol. 23, n°3-4, p. 273-314.