

# Étude de la taille des familles. Réflexions sur l'accès aux données publiques

---

**Bernard AUBRY**

Insee (chargé de mission à la retraite)

*Sur longue période, la taille des familles diminue, comme conséquence naturelle de la baisse de la fécondité. La baisse n'est cependant pas uniforme, elle dépend de différents facteurs, notamment du décalage temporel qui existe entre les composantes de l'indicateur, avec les enfants au numérateur, et les femmes au dénominateur. On constate aussi un agrandissement significatif des familles pour certaines catégories de population aux âges intermédiaires. Comme dans maintes autres situations rencontrées par les chercheurs, l'analyse s'est heurtée aux difficultés d'obtention des données ; c'est pourquoi la seconde partie de la présente communication propose des pistes pour une amélioration significative de l'accès aux données publiques, conformément aux recommandations du Code européen des bonnes pratiques de la statistique.*

## 1. Étude de la taille des familles

### 1.1. Évolutions globales, par tranche d'âge. Les 35-44 ans

S'il existe bien évidemment des liens entre la fécondité – dont on connaît précisément l'évolution sur le long terme – et la taille des familles<sup>7</sup>, les relations ne sont pas aussi nettes qu'on pourrait le penser *a priori*. Dans cette communication, on assimilera, peut-être un peu abusivement, la notion de taille de la famille avec le nombre des enfants. En réalité, la taille de la famille inclut le(s) parent(s), soit une ou plus souvent deux personnes supplémentaires. On calcule donc ici par convention la taille des familles en rapportant le nombre d'enfants au nombre de familles. L'indicateur est ainsi très dépendant des décalages temporels. Un événement marquant sur la fécondité (une forte baisse suivie peu après d'une hausse équivalente par exemple) aura un effet immédiat et progressif sur le numérateur (la baisse sera maximale 9 ans après la chute et une hausse équivalente reprendra symétriquement après). Le dénominateur sera affecté bien plus tard, quand la chute de la fécondité aura atteint les effectifs de femmes en âge de procréer (soit en moyenne 30 ans, mais les premiers effets se ressentiront une dizaine d'années plus tôt). D'autres phénomènes brouillent les liens entre

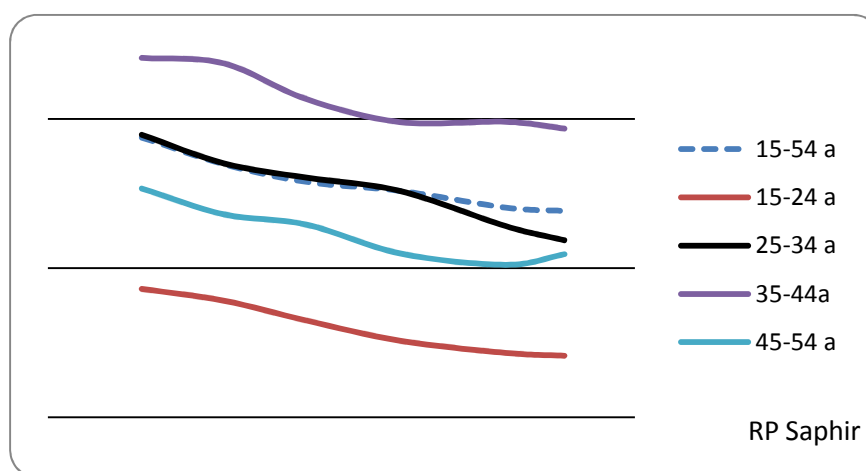
---

<sup>7</sup> Rappelons qu'une famille se compose d'un couple (avec ou sans enfant) ou d'un seul parent avec enfant(s). Pour déterminer le nombre d'enfants, nous avons fixé par convention un âge limite (moins de 18 ans). Au-delà de cet âge, une partie importante des enfants ne vit plus au sein du foyer familial et nous aurions alors des biais perturbant le sens des séries historiques si nous avions pris le seuil de 25 ans. Depuis 1999, la famille intègre les enfants, quel que soit l'âge. Précédemment, il n'y avait famille que si les enfants avaient moins de 25 ans. Cette rupture de définition perturbe les séries brutes. L'une des caractéristiques du fichier Saphir, utilisé dans l'étude, est d'avoir harmonisé les définitions afin de donner du sens aux séries historiques. L'harmonisation a porté notamment sur la catégorie socioprofessionnelle, dont on sait qu'elle a été refondue à compter du RP82.

fécondité et taille des ménages. Parmi les causes perturbatrices, on peut citer les variations de calendrier de la maternité, la recomposition des familles, la proportion de femmes sans descendance et les échanges migratoires. Cette liste n'est pas exhaustive.

Autres conventions : les familles prises en compte dans cette communication correspondent à celles résidant en métropole aux différents recensements de 1968 à 2005 (prise en compte des quatre premières campagnes du nouveau recensement). Les familles monoparentales dont le parent est un homme n'ont pas été introduites dans l'étude. Enfin, par souci de simplification, les familles sont décomposées selon la classe d'âge de la mère (classes de 10 ans, soit quatre classes : 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans).

GRAPHIQUE 1 : ÉVOLUTION DU NOMBRE D'ENFANTS DANS LES FAMILLES, PAR TRANCHE D'ÂGE DE LA MÈRE (1968-2005)



On observe une tendance à la baisse de la taille moyenne des familles, mais celle-ci est nuancée. Depuis 1990, la diminution ne touche de façon très sensible que les 25-34 ans (la baisse s'amplifie même nettement, conséquence sans doute des maternités retardées), elle affecte moins les 15-24, groupe chez qui la baisse avait été plus précoce. Le groupe ayant de loin le plus d'enfants (celui des 35-44 ans) reste assez stable, cependant que le plus âgé (les 45-54 ans) semble engager un mouvement significatif à la hausse. L'indicateur moyen (15-54 ans, en pointillé) a été longtemps très proche de celui du groupe des 25-34 ans, avant que ce dernier ne décroche.

L'indicateur moyen est très affecté par des raisons structurelles. Un indicateur plus sophistiqué sommant les tailles observées à chaque âge de la femme (à l'instar de l'indicateur conjoncturel de fécondité) serait peut-être plus judicieux. Toutefois, du fait qu'il n'a pas été possible d'accéder au fichier *ad hoc* au moment où cette communication a été préparée, ce calcul n'a pas été effectué. Le tableau 1 illustre la complexité de l'indicateur, puisque le numérateur (le nombre d'enfants) et le dénominateur (le nombre de familles) évoluent de façon différente d'une période à l'autre.

Globalement, on observe que le nombre d'enfants vivant en famille a diminué d'environ un demi-million entre 1968 et 2005, cependant que le nombre de familles s'est accru fortement, de 2,5 millions. Les changements intervenus concernent surtout les familles

dont la mère a moins de 25 ans, puisque leur nombre s'est réduit dans le rapport de 1,7 à 1 en près de 40 ans, cependant que le nombre des enfants correspondant s'est réduit de 3,4 à 1, ramenant à 1,4 % leur contribution au total des enfants pris en compte dans l'étude. En toute rigueur, il faudrait aussi prendre en compte les enfants vivant avec une mère de plus de 55 ans (un peu moins de 0,2 enfant par famille ces dernières années, un peu plus de 0,4 enfant en 1968).

**TABLEAU 1 : NOMBRE DE FAMILLES ET NOMBRE D'ENFANTS DANS CES FAMILLES : FEMMES  
ÂGÉES DE 15 À 54 ANS (1968-2005)**

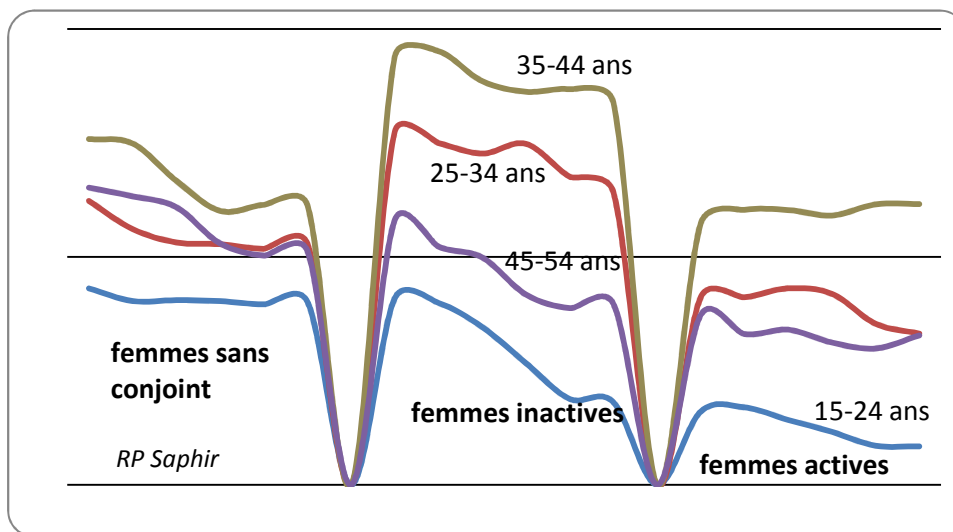
	<b>Effectifs selon l'âge de la mère</b>									
	<b>Valeurs absolues (milliers)</b>					<b>En pourcentages</b>				
	Ensemble	15-24	25-34	35-44	45-54	15-24	25-34	35-44	45-54	Ens.
Enfants - 1968	15 855	741	4 646	7 108	3 360	4,7	29,3	44,8	21,2	100
Enfants - 2005	15 336	218	3 975	7 600	3 542	1,4	25,9	49,6	23,1	100
Familles - 1968	8 449	860	2 452	2 948	2 189	10,2	29,0	34,9	25,9	100
Familles- 2005	10 905	506	3 100	3 835	3 464	4,6	28,4	35,2	31,8	100

*Source* : RP Saphir.

On propose une première partition des familles qui distingue trois groupes, présentés de gauche à droite sur le graphique 2. Le premier groupe est formé des familles monoparentales (rappelons qu'il ne s'agit ici que des femmes seules). Dans cette catégorie, la baisse de la taille des familles s'est interrompue depuis une vingtaine d'années. Les écarts d'une tranche d'âge à l'autre sont relativement faibles et la taille des familles les plus jeunes est restée stationnaire, à un niveau relativement élevé. Le second groupe, au centre, est formé des femmes inactives vivant en couple. Les différences par âge sont alors très grandes. Chez les plus jeunes, la chute a été très marquée, le nombre d'enfants diminuant de moitié depuis 1968 ; il semble toutefois que l'indicateur se soit stabilisé à un niveau bas. Le nombre des enfants vivant à la maison reste élevé pour la tranche supérieure (35-44 ans), avec un rythme à la baisse assez modéré. Chez les femmes de 45 à 54 ans, la baisse a été rapide mais la tendance s'est inversée depuis la fin des années 1980.

Enfin, sur la partie droite du graphique apparaît le cas des femmes vivant en couple et exerçant une activité. Ce n'est pas une surprise de constater que, à tous âges, la taille de la famille est moindre que chez les femmes inactives. En revanche, le maintien sur long terme des niveaux de fécondité est tout à fait remarquable. Cela est vrai notamment chez les femmes de 35 à 44 ans, un peu moins chez les plus jeunes. Tout se passe comme si la baisse de la fécondité ne concernait pas les femmes actives.

GRAPHIQUE 2 : NOMBRE D'ENFANTS PAR FAMILLE, POUR TROIS CATÉGORIES (1968-2005)



## 1.2. La CS de l'homme ou celle de la femme, quelle référence ?

On poursuit l'analyse en intégrant la catégorie socioprofessionnelle. Se pose alors la question de la personne de référence, puisque l'on peut considérer la famille selon différents critères : la CS de la femme, certes, mais aussi celle du conjoint quand la femme vit en couple.

Les résultats relatifs à la taille des familles devront être interprétés dans le contexte d'un bouleversement sociodémographique qui touche toutes les catégories (regroupées en 7 postes). Les cadres et les ouvriers notamment feront l'objet d'un examen plus détaillé par la suite, mais une analyse des employés aurait aussi son intérêt au vu des changements observés dans cette catégorie. Le tableau 2 présente une vision d'ensemble pour cinq groupes de familles observées. Il s'agit, rappelons-le, des seules femmes âgées de 35 à 45 ans, celles-ci formant le groupe le plus important de la population étudiée.

Ce tableau, qui montre combien les indicateurs peuvent être différents selon le point de vue que l'on porte sur la famille, est riche d'informations. Il comprend deux blocs. Celui du haut donne des effectifs (nombre de familles, en milliers), celui du bas la taille des mêmes familles. Pour simplifier, on ne fait figurer que les chiffres des années extrêmes de la période d'étude (1968 et 2005). En colonne, on présente les cinq catégories retenues : les familles désignées selon la CS de l'homme et celle de la femme, la distinction actives-inactives selon la CS de l'homme et enfin la CS de la femme en cas de monoparentalité.

Pour bien comprendre le tableau, on prend l'exemple des agriculteurs, dont la situation est, certes, extrême. Le nombre des familles dont la mère s'est déclarée agricultrice a été divisé par 6 ou presque (de 202 000 à 36 000), celui des hommes par 3 (de 297 000 à 94 000). Quand l'homme est agriculteur, ce même rapport est de 2,2 si la femme se déclare active et de 10 si la femme se déclare inactive.

En ligne figurent des résultats qui permettent de prendre la mesure des écarts à un moment donné entre les différentes catégories de famille. Ainsi, en 1968, le nombre d'enfants

était de 2,68 dans les familles où le père était ouvrier, alors qu’il n’était que de 1,49 quand la mère était ouvrière. Les différences étaient très fortes également entre les femmes inactives (3,15 enfants) et les femmes actives (1,53 enfant seulement), toujours dans le cas où le conjoint est ouvrier. Enfin, les femmes ouvrières avaient de leur côté 2,17 enfants en moyenne quand elles étaient seules à la tête d’une famille.

TABLEAU 2 : QUELQUES INDICATEURS POUR QUATRE CATÉGORIES DE FAMILLE (1968-2005).  
FEMMES DE 35 À 44 ANS

Catégorie socioprofessionnelle (CS de l’homme ou de la femme, selon les cas)	Année	Ensemble des femmes, selon la CS		Femmes inactives (CS de l’homme)	Femmes actives (CS de l’homme)	Femmes sans conjoint (CS de la femme)
		De la femme	De l’homme			
<b>Nombre de familles (milliers)</b>						
Exploitants agricoles	1968	202	297	110	187	5
	2005	36	94	10	84	2
Cadres sup. Prof. libérales	1968	37	237	168	69	4
	2005	366	621	95	526	45
Employés	1968	408	357	227	130	57
	2005	1 555	321	44	278	272
Ouvriers	1968	218	1 144	809	335	28
	2005	339	1 065	191	874	61
<b>Ensemble</b>	<b>1968</b>	<b>3 886</b>	<b>3 473</b>	<b>974</b>	<b>2 499</b>	<b>414</b>
	<b>2005</b>	<b>3 805</b>	<b>3 228</b>	<b>510</b>	<b>2 718</b>	<b>577</b>
<b>Nombre d’enfants par famille</b>						
Exploitants agricoles	1968	2,57	2,59	2,67	2,54	2,51
	2005	2,04	2,06	2,40	2,01	1,76
Cadres sup. et professions libérales	1968	1,80	2,25	2,49	1,66	1,75
	2005	1,76	1,95	2,45	1,86	1,62
Employés	1968	1,52	2,27	2,74	1,45	1,97
	2005	1,86	1,89	2,49	1,80	1,80
Ouvriers	1968	1,49	2,68	3,15	1,53	2,17
	2005	1,89	2,03	2,71	1,88	1,90
Inactifs	1968	2,83	2,33	2,69	1,59	3,10
	2005	2,51	1,82	2,14	1,65	2,49
<b>Ensemble</b>	<b>1968</b>	<b>1,98</b>	<b>2,00</b>	<b>2,59</b>	<b>1,77</b>	<b>1,80</b>
	<b>2005</b>	<b>1,94</b>	<b>1,95</b>	<b>2,52</b>	<b>1,85</b>	<b>1,84</b>

Les graphiques qui suivent vont nous permettre de regarder de plus près les évolutions dans les cas de figure les plus fréquents, en ne considérant toujours que les femmes âgées de 35 à 44 ans.

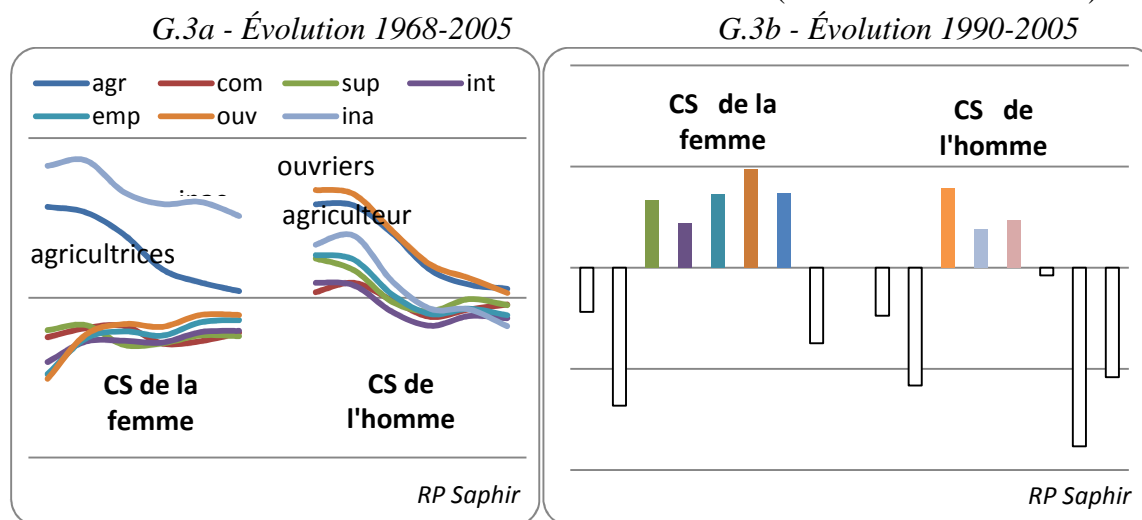
### 1.2.1. Des évolutions divergentes selon le parent de référence

Le graphique 3a reprend l’évolution de la taille des familles selon la CS de la femme (à gauche) et selon la CS de l’homme (à droite), aux différents recensements. S’agissant de la CS de la femme, la caractéristique principale est que cinq des sept catégories sociales sont très proches les unes des autres tant en niveau qu’en évolution : une augmentation certes très faible apparaît même sur le long terme. En revanche, on assiste à une très forte baisse de la taille des familles, tant quand la femme est inactive que lorsqu’elle est agricultrice.

Quand on prend en compte la taille des familles selon la CS de l'homme, l'allure des courbes est notablement différente. Deux d'entre elles se confondent quasiment : celle des ouvriers et celle des agriculteurs. La baisse, qui a été très vive, rapproche sensiblement le comportement de ces deux catégories de celui des cinq autres qui, de leur côté, évoluent parallèlement. L'étiage qui affecte l'année 1990 mériterait d'être analysé de façon plus approfondie.

Le graphique 3b met précisément en relief les tendances qui ont prévalu au cours de la période postérieure à l'année 1990. On maintient la distinction entre les deux points de vue : à gauche, la famille désignée par la CS de la mère, à droite, la famille désignée par celle du père. Globalement, c'est-à-dire pour l'ensemble des familles, on assiste à une baisse de 0,04 point environ. Il n'y a là rien que de très normal, puisque le contenu des deux populations est pratiquement identique (aux familles monoparentales près).

GRAPHIQUE 3 : ÉVOLUTION DU NOMBRE D'ENFANTS PAR CATÉGORIE SOCIOPROFESSIONNELLE, SELON LA CS DE LA FEMME ET CELLE DE L'HOMME (FEMMES DE 35 À 44 ANS)



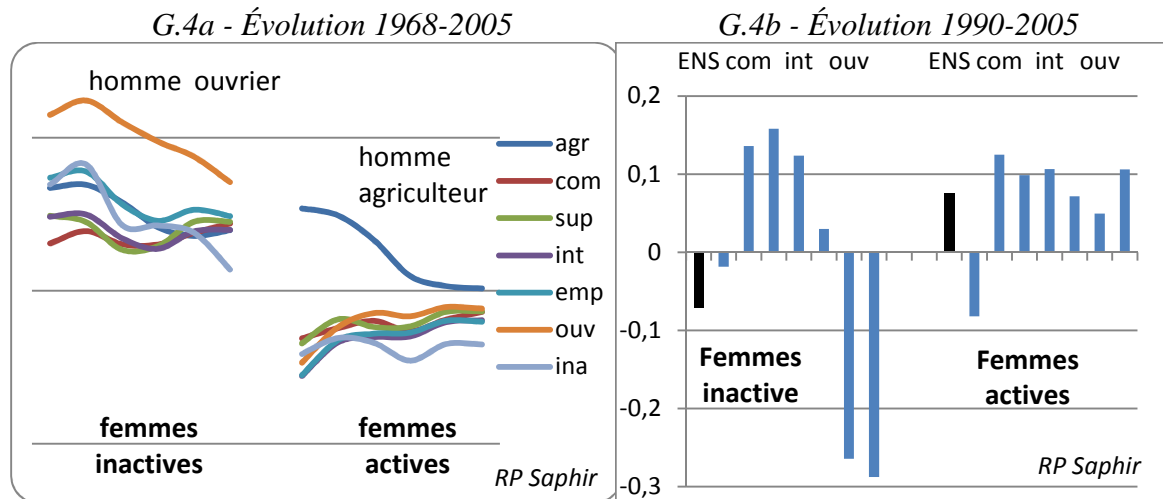
Les deux graphiques ont en commun une baisse sensible de la taille des familles chez les agriculteurs et chez les inactifs. L'intérêt de cette représentation réside dans les différences apparaissant dans les autres groupes : croissance partagée chez les indépendants (commerçants, artisans, chefs d'entreprise), les cadres supérieurs et les professions intermédiaires, divergences d'évolution chez les employés et les ouvriers. Pour cette dernière catégorie, le contraste est même saisissant. Il sera examiné de plus près ultérieurement.

### 1.2.2. Les femmes actives et les femmes inactives

La seconde distinction est faite entre, d'une part, les femmes actives et, d'autre part, les femmes inactives pour les femmes en couple, **selon la catégorie socioprofessionnelle de l'homme**. Comme précédemment, l'une des séries présente un comportement atypique, mais ce n'est pas la même catégorie qui est concernée. Quand la femme est inactive (à gauche), la famille est nettement plus grande quand le conjoint est ouvrier, bien que la tendance à la baisse amorcée au milieu des années 1970 ne se soit pas interrompue, bien au contraire : les femmes d'ouvriers sont encore aujourd'hui celles qui ont le plus d'enfants avec elles. Hormis

le cas des inactifs (quand la conjointe a 40 ans, les hommes inactifs sont peu nombreux), les différentes courbes marquent une nette reprise après avoir connu une baisse assez marquée depuis 1968. Chez les femmes actives, toujours en retrait pour la taille de la famille, ce sont les conjointes d'agriculteurs qui ont enregistré une baisse, sans toutefois rejoindre le bloc formé par les autres groupes. Ceux-ci se caractérisent par une tendance à la hausse de la taille de la famille, faible il est vrai, mais permanente.

GRAPHIQUE 4 : NOMBRE D'ENFANTS PAR FAMILLE SELON LA CS DE L'HOMME : DISTINCTION ENTRE LES FEMMES INACTIVES ET LES FEMMES ACTIVES (FEMMES DE 35 À 54 ANS)



### 1.3. Cadres et ouvriers

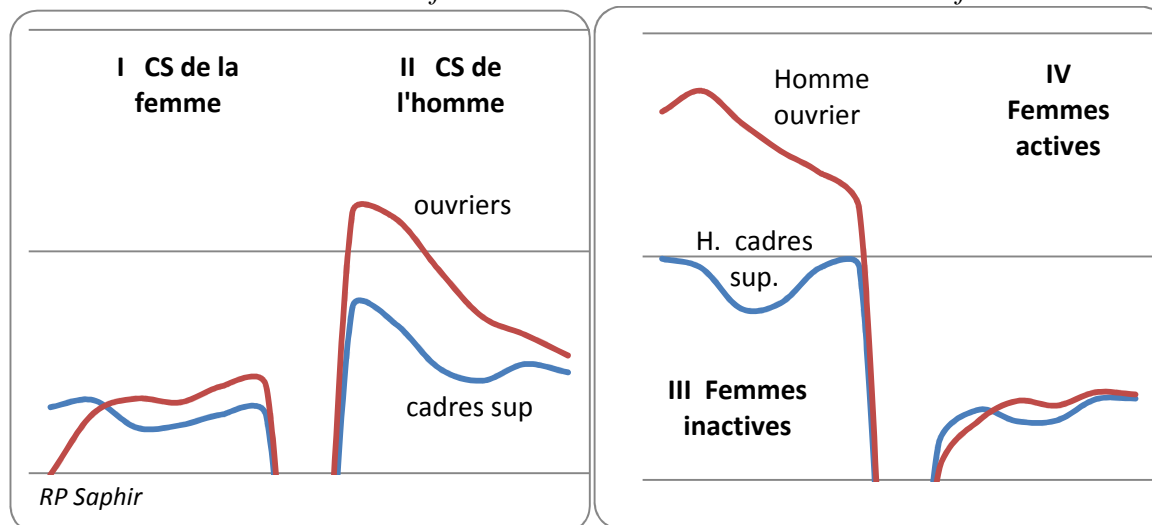
On confronte les évolutions pour les quatre catégories étudiées précédemment (cf. section 1.2.) sous la forme de 4 graphiques explicitant les changements apparus dans ces deux catégories. Ce mode de représentation met en évidence des comportements significativement différents. La forte baisse touchant les familles où l'homme est ouvrier (G.5a à D) a conduit à un rapprochement des courbes, de sorte que, en 2005, les différences entre les deux groupes ne sont plus très grandes. Néanmoins, le nombre d'enfants reste toujours plus grand quand on catégorise la famille par la CS de l'homme plutôt que par celle de la femme.

Une grande analogie apparaît entre les graphiques I (G.5a) et IV (G.5b). Dans les deux cas, les courbes se croisent mais restent toujours peu éloignées l'une de l'autre. Dans le cas I, il s'agit de toutes les femmes ouvrières ou cadres, quelle que soit leur situation par rapport à leur conjoint ; dans le cas IV, il s'agit des femmes actives quelle que soit leur CS propre : elles sont conjointes d'un ouvrier ou d'un cadre. Enfin (G.5b, III), quand elles sont inactives, les femmes d'ouvriers ont toujours beaucoup plus d'enfants avec elles que lorsqu'elles sont conjointes d'un cadre supérieur. Malgré leur rapprochement, les deux courbes restent très éloignées l'une de l'autre. À noter un comportement intéressant : un décrochage est apparu dans les années 1980, mais la remontée qui a suivi a permis à la courbe de retrouver le niveau qui était celui des années 1960.

GRAPHIQUE 5 : ÉVOLUTION DE LA TAILLE DES FAMILLES POUR LES CADRES SUPÉRIEURS ET LES OUVRIERS, SELON LA CS DE L'HOMME ET DE LA FEMME - FEMMES DE 35 À 44 ANS (1968-2005)

G.5a CS de l'homme/CS de la femme

G.5b- Femme active/femme inactive



#### 1.4. De l'indicateur de l'homme à celui de la femme : les ouvriers

On poursuit l'analyse en croisant la catégorie socioprofessionnelle de l'homme avec celle de la femme. On limite notre investigation aux seuls ouvriers.

Le tableau 3 montre l'asymétrie des situations, mais aussi l'ampleur des changements. En 1968, la très grande majorité des ouvriers (70,7 %) avaient une conjointe inactive, contre seulement 18 % en 2005. Aujourd'hui les conjoints d'ouvriers sont majoritairement des employés (48,8 %). Seulement 15,9 % des ouvrières ont un conjoint ouvrier. Les femmes ouvrières, de leur côté, ont le plus souvent un conjoint de la même catégorie socioprofessionnelle, sans qu'aucun grand changement ne soit perceptible depuis 1968 (de 59 % à 50 % pour les conjoints ouvriers).

Au regard de ce contexte structurel, l'évolution de la taille des familles prend alors une autre signification. Ainsi, chez les hommes, la taille des familles baisse en près de 40 ans de 0,65 point (de 2,68 enfants à 2,03), plus que le sous-groupe le plus important que forment les femmes inactives (- 0,44 point), dont le poids passe de 70,7 % à 18,0 %. Les autres catégories marquent en général des évolutions fortement positives—: ainsi, + 0,51 point pour les employées, + 0,31 pour les professions intermédiaires. Chez les femmes, l'augmentation moyenne de la taille est de 0,5 point pour l'ensemble des conjoints. Cette augmentation touche toutes les catégories, y compris les familles où l'homme se déclare agriculteur ou commerçant. La taille des familles s'accroît même de façon très significative dans le cas des femmes ayant un conjoint employé (+ 0,58) ou appartenant à une profession intermédiaire (+ 0,57 point).



TABLEAU 3 : CHEZ LES OUVRIERS, DE LA CS DE L'HOMME À CELLE DE LA FEMME - FEMMES DE 35 À 44 ANS (EN 1968, 1990 ET 2005)

	Nombre de familles						Taille des familles		
	En milliers			En pourcentages			1968	1990	2005
	1968	1990	2005	1968	1990	2005			
<b>Homme ouvrier - CS de la femme</b>									
Ensemble	1 144	1 199	1 065	100,0	100,0	100,0	2,68	2,21	2,03
Agricultrices	13	7	4	1,2	0,6	0,3	2,66	2,07	1,91
Commerçantes	18	20	19	1,6	1,6	1,7	1,82	1,80	1,79
Cadres sup.	2	13	26	0,2	1,0	2,4	1,48	1,60	1,66
Prof. interm.	22	87	137	2,0	7,3	12,9	1,46	1,74	1,77
Employées	151	439	520	13,2	36,6	48,8	1,51	1,85	1,92
Ouvrières	128	240	169	11,2	20,0	15,9	1,41	1,84	1,91
Inactives	809	394	191	70,7	32,8	18,0	3,15	2,97	2,71
<b>Femme ouvrière – CS de l'homme</b>									
Ensemble	218	425	339	100,0	100,0	100,0	1,49	1,82	1,89
Agriculteurs	2	5	7	0,8	1,3	2,2	1,54	1,81	1,94
Commerçants	7	21	16	3,3	4,9	4,6	1,39	1,76	1,89
Cadres sup.	3	11	12	1,2	2,6	3,5	1,26	1,69	1,79
Prof. interm.	20	56	42	9,2	13,2	12,4	1,28	1,70	1,85
Employés	24	29	23	10,9	6,7	6,7	1,32	1,81	1,90
Ouvriers	128	240	169	58,8	56,5	49,7	1,41	1,84	1,91
Inactifs	6	9	10	2,8	2,2	2,9	1,66	1,74	1,82

Source : RP Saphir.

### 1.5. Pour aller plus loin

Cette étude a été réalisée à partir des données contenues dans un seul tableau obtenu par exploitation du fichier Saphir<sup>8</sup>. Bien que ce tableau ne croise qu'un petit nombre de variables (CS de l'homme, de la femme, du département pour six recensements et six classes d'âge), sa taille est cependant conséquente (de l'ordre d'un million de lignes). On aurait certes pu poursuivre l'investigation en prenant en compte le département, ce qui aurait apporté une certaine image de la transformation du territoire, mais on aurait aimé aller plus loin. Disposer par exemple de variables telles que la commune de résidence et la taille du logement (nombre de pièces) aurait permis de faire des investigations adaptées à l'étude de la périurbanisation. On aurait aimé également relier l'évolution de la taille des familles avec le diplôme de la mère, puisque ce facteur est réputé fortement corrélé avec la fécondité ; on aurait souhaité savoir ce qu'il en est de l'origine des parents nés hors métropole, l'information étant alors complétée par l'ancienneté de présence dans l'Hexagone, puisque ce facteur n'est pas sans lien avec la distribution de la taille de la famille (qu'en est-il de la fréquence des familles nombreuses; ? Quelle est alors la contribution des immigrés ?). Une dimension supplémentaire aurait pu être apportée à l'étude si les résultats observés avaient été mis en relation avec les événements qui ont traversé la société au cours des quarante dernières années. Qui cherche à mesurer les transformations de la société, qui cherche à évaluer quantitativement les décisions politiques doit pouvoir accéder aux données adéquates sans

<sup>8</sup> Saphir : Système d'analyse de la population par l'historique des recensements, mis au point à l'Insee-Alsace.

que la recherche de l'information nécessaire et existante ne se transforme en une course d'obstacles inutile et coûteuse. En l'occurrence, l'analyse s'est trouvée entravée par l'impossibilité d'accéder aux variables individuelles. En effet, seul un accès aux données de base permet de satisfaire l'attente des utilisateurs exigeants.

## 2. Réflexion sur l'accès aux données publiques

Nombre de chercheurs se plaignent des difficultés qu'ils rencontrent dans la quête des informations statistiques dont ils ont besoin pour leurs travaux d'investigation. Cette critique n'est pas nouvelle et je l'entendais déjà au début de ma carrière à l'Insee. Elle ne concerne d'ailleurs pas seulement les chercheurs mais également tous ceux qui ont à traiter l'information statistique publique. Avant l'irruption de l'informatique, le traitement et le stockage des données limitaient évidemment les possibilités de diffusion, mais ce n'est plus le cas aujourd'hui avec la généralisation de l'internet. La démarche qui a conduit, dès la fin des années 1980, à la création d'un fichier historique des recensements (Saphir) à la DR de l'Insee-Alsace, fichier utilisé dans cette communication, n'était pas étrangère à ce souci de répondre plus efficacement aux attentes des utilisateurs en regroupant en un seul le contenu de plusieurs fichiers diversement organisés. Une telle base historique permet de s'assurer d'une bonne cohérence d'ensemble. Elle permet en outre d'augmenter de façon significative la productivité, en libérant les agents de tâches préparatoires longues et fastidieuses. Néanmoins, une telle démarche ne semble pas avoir suscité, au sein de l'institution, l'intérêt qu'on aurait pu attendre, puisqu'elle n'a jamais fait l'objet d'une évaluation, ni même d'une discussion collective. Le fichier a été largement boudé à l'intérieur même de l'Insee. Faut-il donc s'étonner si un outil non promu au sein de l'institution le soit moins encore à l'extérieur ? Le rapport du Conseil d'analyse économique (CAE) intitulé *Immigration, qualifications et marché du travail*, remis le 27 mai 2009 au ministre de l'Immigration, illustre bien le défaut de diffusion. En effet, ce rapport **interrompait en 1990** la présentation de séries historiques sur les concentrations d'immigrés et notait, pour le regretter, l'absence de données plus récentes<sup>9</sup>. Et pourtant, au même moment, la revue *Aspenia*, puis la revue *Commentaire* publiaient des séries historiques de contenu voisin jusqu'en 2005<sup>10</sup>. Celles-ci avaient été obtenues par simple adjonction au fichier Saphir d'une sélection de données issues des quatre premières campagnes des enquêtes de recensement. C'est ce même fichier qui a pu être utilisé pour l'examen de l'évolution de la taille des familles. Peut-être un jour le fichier Saphir sera-t-il officiellement parachevé et contiendra-t-il l'intégralité du recensement 2006<sup>11</sup>, ce qui

---

<sup>9</sup> On lit, en renvoi, page 224 du rapport (*Immigration, qualifications et marché du travail*, La Documentation française. Paris, 2009 – ISBN : 978-2-11-007726-4) : « Pour le recensement de 1999, il est impossible d'accéder à un certain nombre d'informations à un niveau géographique détaillé, qui est celui dont nous avons besoin pour analyser convenablement les concentrations d'immigrants. Cette difficulté n'existe pas pour les recensements antérieurs ».

<sup>10</sup> *Aspenia*, n° 44 (« I giovani stranieri di Francia » – Bernard Aubry et Michèle Tribalat), *Commentaire*, n° 126, été 2009 (« Les jeunes d'origine française » – mêmes auteurs). L'Insee, à qui le projet de texte avait été proposé fin 2008, n'avait pas souhaité le publier.

<sup>11</sup> On peut accéder en ligne à une base de données individuelles intitulée : « Données harmonisées des recensements de 1968 à 1999 ». Quoiqu'il ne soit fait nullement mention à l'acronyme, on a tout lieu de penser

permettrait de suivre la démographie française de 1962 à 2006 ? Pourquoi faire attendre aussi longtemps les utilisateurs ? L'article 15 du Code européen des bonnes pratiques de la statistique préconise l'accessibilité et la clarté aux données publiques. De son côté, l'article 14 porte sur la cohérence et la comparabilité des données.

On se réjouit de la mise en ligne de fichiers individuels, seuls en mesure de satisfaire les besoins des chercheurs, tout en regrettant toutefois qu'ils n'autorisent encore que peu d'exploitations au niveau infrarégional<sup>12</sup>. Mais le problème est aussi lié à la comparabilité des résultats dans le temps. S'agissant du recensement, il est certain que, du fait de la rupture engendrée par le caractère novateur de la méthodologie du nouveau recensement, une certaine prudence s'imposait dans la mise à disposition des données séquentielles. Il reste que les expériences que nous avons faites (cf. la communication sur la taille des familles par catégories socioprofessionnelle, et les études publiées sur les enfants immigrés) plaident en faveur d'une bonne continuité des séries. Probablement pas de toutes. Cependant, on n'imagine pas qu'un chercheur n'ait pas le souci de comparer les chiffres les plus récents, déjà disponibles, à ceux du passé, encore peu accessibles. Plutôt que de laisser tout un chacun « bricoler » à sa façon des séries, avec tous les risques inhérents à de telles pratiques, ne vaut-il mieux confier cette tâche d'harmonisation aux spécialistes de l'Insee qui sont rompus aux arcanes des méthodes de redressements complexes ? Au demeurant, certains chercheurs regrettent que les méthodes de traitement d'enquêtes soient opaques, alors même que l'article 8 du même Code européen évoque la transparence des procédures statistiques. Plus généralement, au-delà du souhait de la transparence, il n'est pas inutile de rappeler la nécessité d'une pédagogie permettant une meilleure compréhension entre les producteurs et les consommateurs de statistique.

Quelques pistes concernant la mise à disposition des données :

– engager une coopération entre l'Insee et les chercheurs pour faciliter la **capitalisation** des travaux d'harmonisation. L'Insee pourrait se charger d'une harmonisation minimale (type Saphir) pour les recensements et certaines enquêtes périodiques, telles celles sur l'emploi et le logement. Pour le reste, les universitaires pourraient être mis à contribution chacun dans leur domaine, l'Insee apportant son assistance et se chargeant de capitaliser les investissements réalisés. Ainsi, par exemple, les séries historiques sur le temps de travail harmonisées par un chercheur en démographie au prix d'un traitement lourd des différentes bases annuelles mises à sa disposition *via* le centre Quetelet pourraient être d'une grande utilité pour d'autres chercheurs<sup>13</sup>. D'une façon générale, tout investissement universitaire qui améliore les bases de données ou les rend plus accessibles pourrait tomber dans le domaine public et, par là même, contribuer à l'extension de l'offre statistique ;

---

qu'il s'agit d'une extraction du fichier Saphir. Fortement amputé, le fichier actuellement disponible sur internet ne peut guère satisfaire les attentes des chercheurs.

<sup>12</sup> Il en est ainsi des études sur la périurbanisation, dans la mesure où le contenu des données individuelles mises en ligne au niveau communal est pauvre.

<sup>13</sup> Je pense notamment à l'expérience de Ceren Inan (Iedub Bordeaux) qui, dans le cadre de sa thèse, a engagé un examen approfondi sur le temps de travail à travers une analyse spécifique des enquêtes emploi. Des séries longues et cohérentes sur le temps de travail, dûment validées par l'Insee, pourraient à n'en point douter être utiles aux chercheurs en économie.

– rendre l’Insee plus réceptif aux remarques, aux critiques faites sur la qualité des résultats, leur fiabilité, mais aussi aux suggestions portant sur le contenu des enquêtes. Les courriers restent trop souvent lettre morte. Les doléances faites à l’Institut de la statistique ne pourraient-elles pas être soumises simultanément au Conseil national de l’information statistique (CNIS) ? Cette procédure aurait un double avantage : sensibiliser les membres du Comité aux difficultés rencontrées par les utilisateurs et inciter l’Insee à donner une réponse aux questions posées ;

– mettre en place un Comité de chercheurs et d’utilisateurs amenés à consulter les sites des autres instituts nationaux de statistique. Chacun, dans son domaine, apporterait une pierre à l’édifice puisqu’il serait invité à comparer méthodiquement le contenu, la qualité et les procédures de mise à disposition des données publiques par les services étrangers homologues de l’Insee. Les résultats de ces investigations seraient eux-mêmes organisés et mis en ligne. La publicité ainsi faite pourrait favoriser l’émulation entre les différents services publics, en un temps où chacun sait l’importance qu’ont pris les chiffres dans la construction européenne. De son côté, la communauté des chercheurs serait partie prenante à l’amélioration de la qualité du patrimoine statistique. Elle se trouverait notamment informée par elle-même de la façon dont les différents offices de statistique traitent les demandes portant sur les données dites sensibles (cryptage, interrogation à distance, etc.), puisqu’il s’agit là d’un problème délicat auquel les responsables des services de diffusion sont confrontés ;

– les universitaires eux-mêmes, de leur côté, font-ils ce qu’il faut pour améliorer la situation ? Ne sont-ils pas résignés face à une institution qui leur apparaît souvent comme une « forteresse » ? Par ailleurs, chacun ne nourrit-il pas le secret espoir d’obtenir un sésame auprès d’un responsable compréhensif<sup>14</sup> ? Pourtant, on rappelle que la statistique est un bien public (« *res nullius* »), que les chiffres n’appartiennent à personne et que les responsables de l’Insee ont la charge de les traiter de façon à les rendre utiles à tous. S’il apparaît que ce n’est pas le cas, les citoyens, en l’occurrence les chercheurs, pourraient, et même devraient alerter les pouvoirs publics et même l’opinion. Personne, et surtout pas l’Insee, n’a à gagner, à un « *affaiblissement*<sup>15</sup> » de la statistique publique qui conduirait très vite à l’effacement de l’institution. Tôt ou tard, la collecte et le traitement des données passeraient dans le domaine marchand et seraient conditionnés aux besoins du moment. Est-ce bien cela que nous voulons ?

---

<sup>14</sup> La « forteresse » s’ouvre parfois, comme en témoigne l’interview de Rahsaan Maxwell paru dans *Le Monde* du 2 février 2010. Cet universitaire américain a pu travailler au sein de l’Insee et accéder ainsi à des fichiers sensibles (sur l’immigration).

<sup>15</sup> La formule est empruntée à Claude Thélot (préface de l’ouvrage *Échantillonnage et méthodes d’enquêtes*, ouvrage collectif publié sous la direction de Pascal Ardilly – Dunod, 2004).