

CODER LA PROFESSION : NOUVELLES PROCEDURES, VIEUX PROBLEMES

A. CHENU () et F. GUGLIELMETTI (**)*

(*) CREST

(**) INSEE - Unité "Méthodes Statistiques"

La profession des individus et leur situation par rapport au marché de l'emploi présentent de multiples facettes (métier, classification professionnelle, taille de l'établissement...). Il n'est donc pas étonnant que le codage statistique de la profession requière un questionnement lui-même multiple et prenne des formes différentes selon qu'il concerne des salariés ou des non-salariés, les salariés d'une petite entreprise ou ceux de la fonction publique, etc.

Le nombre et l'ordre des questions posées en vue du codage de la profession ou de la catégorie socioprofessionnelle varient d'un dispositif statistique à un autre. La diversité des dispositifs est grande. À son niveau à deux chiffres, la catégorie socioprofessionnelle peut être renseignée aussi bien dans des fichiers de DADS (déclarations annuelles de données sociales) que dans des données d'état civil, dans des recensements de la population que dans diverses enquêtes Insee auprès des ménages ou des employeurs, dans des sondages d'opinion à caractère politique ou commercial aussi bien que dans des documents administratifs de toute sorte. Ni le diplôme, ni le revenu, ni aucun autre indicateur de statut social ne présentent une telle « portabilité », ne circulent aussi facilement d'une sphère de description du monde social à une autre.

CS et PCS

Entrée en vigueur en 1982, la nomenclature des professions et catégories socioprofessionnelles (PCS) comporte, dans son niveau le plus détaillé (code à quatre chiffres), 455 positions élémentaires, correspondant à des ensembles homogènes de professions. Le nombre de positions passe à 553 dans la version rénovée (PCS 2000).

Les deux premiers chiffres du code définissent la catégorie socioprofessionnelle (CS). Détaillée en 42 postes, elle peut être utilisée sous une forme regroupée en 24 postes pour les publications courantes. Le niveau le plus agrégé (premier chiffre du code) correspond aux grands groupes socioprofessionnels (8 postes).

Dans le langage à première vue commun des catégories socioprofessionnelles, les mêmes mots ne recouvrent pas toujours les mêmes objets. Les spécialistes de la mortalité différentielle savent par exemple qu'il y a bien plus d'*employés* au sens de l'état civil qu'au sens du recensement. Au sein d'un même dispositif, des changements apparemment minimes dans l'ordre des questions posées ou dans les procédures de traitement de l'information créent parfois de fortes discontinuités dans les séries. On sait qu'à une question sur la classification au sens des conventions collectives, la modalité « employé » est plus souvent choisie si elle vient avant l'item « cadre » que si elle lui fait suite, parce que dans le premier cas elle est prise comme synonyme de salarié, etc.

On se propose de présenter ici quelques observations permettant de caractériser la robustesse du codage socioprofessionnel tel qu'il est pratiqué au recensement et à l'enquête Emploi. Ces deux dispositifs vont connaître dans un proche avenir de substantielles réformes qui auront d'inévitables conséquences sur les chaînes de collecte et de traitement de la profession. On vise aussi à apporter une petite contribution aux vastes interrogations que suscite le devenir des nomenclatures sociales et professionnelles, dans un contexte marqué par des changements dans trois domaines interdépendants : demande sociale (administrations, presse, entreprises, milieux scientifiques...) en matière de statistiques sur les professions, l'emploi, la stratification sociale, technologie des enquêtes statistiques, intégration statistique internationale.

Un contexte changeant

Demande sociale : les « professions et catégories socioprofessionnelles » sont-elles toujours intéressantes ? Sait-on ce qu'on mesure lorsqu'on renseigne la CS ? La profession a-t-elle la même valeur aujourd'hui et hier comme indicateur de position sociale ?

Technologie des enquêtes statistiques : avec l'extension de la collecte assistée par ordinateur et la facilité croissante du traitement automatique des libellés en clair recueillis en réponse à des questions ouvertes, quelles modifications peut-on (ou doit-on), dans tel ou tel dispositif, apporter à la chaîne de collecte et de traitement qui aboutit au codage de la CS et/ou de la PCS ?

Intégration statistique internationale : le particularisme des catégories socioprofessionnelles à la française les voue-t-il à une disparition à terme ? Quels arguments justifieraient le maintien de leurs spécificités, ou leur accès au statut de modèle, dans le cadre d'une Europe statistique tendant à promouvoir successivement une standardisation des nomenclatures sectorielles, professionnelles, sociales ?

On se limite ici à l'examen partiel du deuxième de ces trois volets.

1. Les catégories socioprofessionnelles à l'épreuve de la répétition¹

Depuis une vingtaine d'années, différents travaux ont mis en lumière la variabilité des déclarations qu'une même personne peut fournir à propos d'une situation professionnelle inchangée, et celle des codages qu'un organisme statistique peut réaliser, au long de deux chaînes de traitement différentes, à partir d'une même déclaration de profession [Insee 1981, Merllié 1990, Kramarz 1991, Rouault 1995, Chenu 1997, Héran 1997, Guglielmetti 2000]. Face à cette variabilité, ils ont développé une approche centrée sur la recherche des causes de variation, celle-ci étant considérée comme du signal et non comme du bruit, comme l'expression d'aspects auparavant cachés des situations de travail et d'emploi plutôt que comme du flou indésirable.

La réflexion amorcée ici s'appuie sur deux rapprochements. D'abord celui du codage de la PCS, pour les mêmes personnes interrogées deux fois, au recensement (codage semi-automatique selon la procédure Colibri²) et à l'enquête Emploi de 1982 (codage

¹ Cf. l'article de même titre de D. Merllié (1990).

² Colibri : COdage en LIgne des Bulletins du Recensement Individuel

« manuel » à l'aide d'une documentation principalement constituée d'un guide et d'un index alphabétique sur papier) [Chenu 1997]. Ensuite celui du codage de la PCS et des variables connexes au test du recensement de 1997 à partir des mêmes bulletins traités de manière semi-automatique d'une part, et selon une procédure de contrôle principalement « manuelle » d'autre part³ [Guglielmetti 2000].

Au rapprochement de 1982, les divergences entre les deux chiffrements de la PCS résultent tantôt des différences dans la nature des déclarations que les personnes ont fournies dans le cadre du recensement et de l'enquête Emploi (« flou de déclaration »), tantôt de celles dans le codage effectué par l'Insee (« flou de classement »), tantôt d'un cumul des deux. Elles peuvent être particulièrement fréquentes parce que 1982 est l'année où l'on met en oeuvre pour la première fois une toute nouvelle nomenclature socioprofessionnelle [Desrosières et Thévenot 1989]. Par ailleurs les deux procédures ne sont pas de même nature : le questionnaire de l'enquête Emploi est administré par un enquêteur, celui du recensement est déposé puis récupéré par un agent recenseur ; le recensement « général » de la population est une opération qui sollicite avant tout le civisme des répondants [Kramarz 1991], tandis que le contexte d'une enquête focalisée sur l'*emploi* invite à détailler la description des situations de travail ou de chômage.

³ Un des objectifs du test du recensement est d'évaluer et d'améliorer la chaîne de codage de la PCS. La mise en oeuvre de l'application Sicore-PCS (Sicore : Système Informatique de COdage par REcognition) permet de coder environ sept bulletins sur dix. Les bulletins restants font l'objet d'une « reprise » manuelle (avec la documentation papier), reprise elle-même informatiquement assistée par Sicore. Tous les bulletins codés automatiquement font l'objet d'un codage de contrôle manuel. Les autres sont codés une seconde fois en reprise. L'analyse des écarts entre les deux codages conduit à des modifications de la base de règles Sicore et de l'organisation de la reprise manuelle.

³ Les différences résultent de mécanismes complexes qui mériteraient une analyse approfondie. A chaque divergence, il faut répondre à toute une série de questions : existe-t-il une règle de codage précise ? Si oui, est-elle la même dans les tables du codage automatique et dans les instructions papier ? Si elle est unique, pourquoi n'est-elle pas appliquée dans un des cas : non connue, jugée erronée ? S'il n'existe pas de consigne précise, quels critères implicites ont été mis en oeuvre dans les deux codages ? Résultent-ils de la formation, de consignes générales locales, de l'initiative individuelle ?

Au test de 1997, les différences résultent uniquement de variations dans les procédures de codage à partir d'un même bulletin⁴.

Globalement, parmi les actifs occupés au sens du recensement, le taux de divergence dans le codage de la PCS à quatre chiffres est de 43 % au rapprochement de 1982, et 29 % à celui de 1997.

Taux de divergence entre les deux codages socioprofessionnels

- au recensement et à l'enquête Emploi de 1982

Groupe socioprofessionnel au recensement de 1982	Classement à l'enquête Emploi de 1982				total	score D*
	(p ₀) groupe différent	(p ₁) même groupe catégorie différente	(p ₂) même catégorie profession différente	(p ₃) même profession		
1. agriculteurs	5	24	28	43	100	70
2. indépendants	13	7	17	63	100	77
3. cadres	24	10	19	47	100	62
4. professions intermédiaires	19	7	21	53	100	69
5. employés	13	5	19	63	100	77
6. ouvriers	9	28	16	47	100	67
Ensemble	13	15	19	53	100	71

⁴ Les différences résultent de mécanismes complexes qui mériteraient une analyse approfondie. A chaque divergence, il faut répondre à toute une série de questions : existe-t-il une règle de codage précise ? Si oui, est-elle la même dans les tables du codage automatique et dans les instructions papier ? Si elle est unique, pourquoi n'est-elle pas appliquée dans un des cas : non connue, jugée erronée ? S'il n'existe pas de consigne précise, quels critères implicites ont été mis en oeuvre dans les deux codages ? Résultent-ils de la formation, de consignes générales locales, de l'initiative individuelle ?

- au traitement semi-automatique et au contrôle manuel du test de recensement de 1997

Groupe socioprofessionnel au codage semi-automatique	Classement au codage de contrôle				total	score D*
	(p ₀) groupe différent	(p ₁) même groupe catégorie différente	(p ₂) même catégorie profession différente	(p ₃) même profession		
1. agriculteurs	3	24	12	61	100	77
2. indépendants	8	6	12	74	100	84
3. cadres	11	10	18	61	100	76
4. professions intermédiaires	12	7	12	69	100	79
5. employés	8	4	11	77	100	86
6. ouvriers	5	12	12	71	100	83
Ensemble	8	8	13	71	100	82

*Voir l'encadré.

Conformément aux attentes, le cumul d'un flou de déclaration et d'un flou de classement produit dans toutes les catégories des écarts plus forts en 1982. Mais la similitude des profils aux deux rapprochements est frappante. Les principales conclusions du rapprochement de 1982 [Chenu 1997] valent en 1997. En termes de « descriptibilité » statistique, les actifs occupés se répartissent en trois grands ensembles :

- Le premier, qui comprend les agriculteurs et les ouvriers, se définit par la sûreté de l'appartenance au groupement à un chiffre et l'incertitude des divisions internes (notamment au niveau des catégories à deux chiffres).
- Le deuxième, composé des employés et des indépendants, se prête à une bonne description en gros comme en détail (le niveau qui « marche » le mieux est celui des catégories à deux chiffres).
- Le troisième, avec les cadres et les professions intermédiaires, se caractérise par beaucoup de flottement au niveau le plus agrégé et des taux de divergence moyens (ou médiocres dans le cas des cadres en 1997) dans les classements de détail.

Un score global de « descriptibilité » statistique peut mesurer l'aptitude moyenne des doubles processus de codage à opérer des classements qui convergent à la fois en gros et en détail (voir l'encadré). En 1997 comme en 1982, il atteint ses valeurs les plus élevées chez les employés, les indépendants, les agriculteurs. Mais cet indicateur gomme la configuration différenciée des écarts de codage qui, très schématiquement, oppose les agriculteurs et les ouvriers, vastes ensembles aux divisions internes peu robustes, et les autres regroupements, dont les contours

globaux sont moins nets mais dont la structuration interne en professions ou en catégories socioprofessionnelles à deux chiffres est d'une meilleure lisibilité.

Un score de descriptibilité

Le croisement du codage de la PCS pour les mêmes personnes caractérisées à deux reprises permet de situer les catégories au long d'une échelle de « descriptibilité » à quatre positions.

Lorsque même le repérage le plus grossier, celui parmi l'un des six groupes socioprofessionnels, diffère aux deux sources (rapprochement enquête Emploi-recensement de 1982) ou au terme des deux procédures de codage des mêmes bulletins (test du recensement de 1997), la descriptibilité des personnes concernées est au niveau zéro. On se situe au niveau 1 lorsque la catégorie à deux chiffres diffère mais que le premier chiffre est le même, au niveau 2 lorsque la profession à quatre chiffres diffère mais que la catégorie est la même dans les deux dispositifs, et enfin au niveau 3, le plus élevé possible, lorsque les deux codages à quatre chiffres coïncident.

La série des quatre pourcentages p_0 à p_3 décrivant la composition d'une population d'effectif n au regard de cette échelle constitue le profil de descriptibilité de cette population dans la nomenclature des PCS. Sous une hypothèse rustique d'équidistance entre les barreaux successifs de l'échelle, on peut définir un score de descriptibilité socioprofessionnelle D , égal à la moyenne des scores individuels :

$$D = (p_1 + 2p_2 + 3p_3) / 3$$

D prend la valeur minimale 0 si aucun individu n'est classé dans le même groupe socioprofessionnel dans les deux dispositifs, et la valeur maximale 100 si tous les codages coïncident au niveau le plus détaillé de la nomenclature.

Le fait que les scores de descriptibilité soient médiocres dans le cas des cadres et membres des professions intellectuelles supérieures est surprenant. C'est au sein de cette catégorie que les emplois sont le plus fortement institutionnalisés. Les uns le sont sur le mode des *corps* de la fonction publique, d'autres se rapprochent des *professions* au sens anglo-saxon du terme (formation de niveau supérieur contrôlée par la profession elle-même, monopole d'exercice garanti légalement, déontologie spécifique). Dans tous les cas, les contours des professions ont en principe une grande visibilité sociale. La nomenclature est-elle trop détaillée ? Les procédures de questionnement et de codage sont-elles inadaptées ? Toujours est-il que la forte variabilité des codages requiert ici des explications particulières.

Les profils de descriptibilité des employés et des cadres varient beaucoup selon le sexe : la catégorisation des hommes comme cadres et des femmes comme employées, situations professionnelles assez typiques, est robuste, alors que celle des hommes

employés et des femmes cadres est plutôt incertaine (chiffres non reproduits ici ; pour 1982, voir Chenu 1997).

L'observation des « échanges » entre les groupes, respectivement au recensement et à l'enquête Emploi, confirme la robustesse des phénomènes décrits ci-dessus.

Classement des actifs par groupe socioprofessionnel

- au recensement et à l'enquête Emploi de 1982

Recensement	Enquête Emploi						Ensemble
	agriculteurs	indépendants	cadres	prof.interm.	employés	ouvriers	
agriculteurs	65	1	0	0	0	2	68
indépendants	0	74	4	3	3	1	85
cadres	0	1	67	17	1	1	87
prof. intermédiaires	0	2	11	142	13	8	176
employés	0	1	3	17	222	13	256
ouvriers	1	1	0	12	16	298	328
Ensemble	66	80	85	191	255	323	1000

- au traitement semi-automatique et au contrôle manuel du test de recensement de 1997

Codage semi-automatique	Codage de contrôle						Ensemble
	agriculteurs	indépendants	cadres	prof.interm.	employés	ouvriers	
agriculteurs	66	0	0	0	0	2	68
indépendants	0	78	3	2	1	1	85
cadres	0	1	77	6	2	1	87
prof. intermédiaires	0	2	7	155	8	4	176
employés	0	0	2	8	236	10	256
ouvriers	0	1	0	3	12	312	328
Ensemble	66	82	89	174	259	330	1000

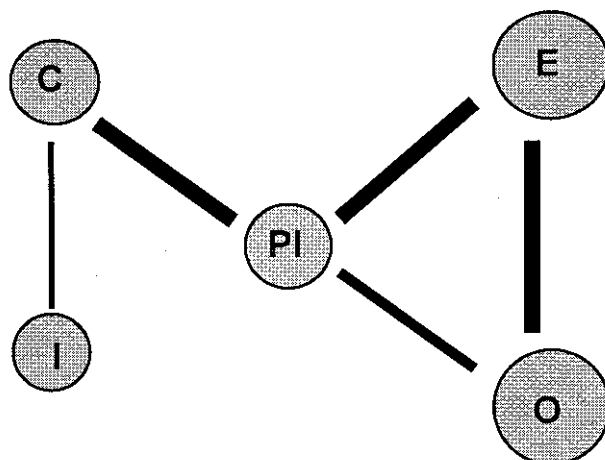
Dans les deux cas, ce sont les mêmes groupes qui échangent entre eux : cadres et professions intermédiaires, professions intermédiaires et employés, employés et ouvriers. Le rapprochement avec la mesure de la qualité au recensement de 1982 confirme cette très grande permanence des flottements entre les groupes [Guglielmetti 2000, p.33].

La « dérive » sur les effectifs globaux des groupes, à l'expertise du test du recensement de 1997 comme à l'appariement recensement-enquête Emploi de 1982, est dans l'ensemble assez faible. Cependant nombre de cadres (et aussi d'employés et

d'ouvriers) au sens du recensement sont classés dans les professions intermédiaires à l'enquête Emploi. Le groupe des professions intermédiaires est aussi celui qui, au test du recensement de 1997, donne lieu aux reprises les plus fréquentes (24% de libellés non reconnus). Le questionnement approfondi de l'enquête Emploi repère un plus grand nombre de situations « intermédiaires », celui du recensement réserve une plus large place aux descriptions les plus « typifiées », il durcit les contrastes.

Le schéma ci-après vaut pour le recensement comme pour l'enquête Emploi :

Attractions entre catégories socioprofessionnelles du fait du double codage



Lecture :

I	Indépendants
C	Cadres
PI	Professions intermédiaires
E	Employés
O	Ouvriers

L'épaisseur du trait est proportionnelle au volume des échanges.
La surface des cercles est proportionnelle aux effectifs.

2. Les libellés de profession en clair

La question ouverte sur la profession était, jusqu'au recensement de 1962, l'unique question (outre le secteur d'activité économique) posée en vue du codage du métier et de la catégorie socioprofessionnelle, et elle est toujours la première question posée dans la partie emploi du bulletin du recensement et dans les enquêtes auprès des ménages.

Si, au recensement, on se contentait de la déclaration de la profession en clair pour coder dans la nomenclature des PCS, on obtiendrait 28% de codes « erronés », dont 10% au niveau du groupe socioprofessionnel (*source : test du recensement de 1997*). Ce résultat est susceptible d'une interprétation contradictoire : le questionnement sur les « variables annexes » (statut, position professionnelle, fonction, activité économique de l'établissement employeur...), qui sont sollicitées à des degrés divers (le statut l'est presque toujours) est d'une grande importance pour le codage de la profession. Mais inversement, sept fois sur dix, les questions complémentaires défilent *in fine* pour rien, ne fournissant qu'une information redondante. Le codage de la profession s'appuie avant tout sur un intitulé en clair.

On a vu que près d'une personne sur deux interrogée sur sa profession respectivement au recensement et à l'enquête emploi est classée à ces deux sources sous deux rubriques différentes. Cette divergence peut résulter pour partie de différences dans la façon dont est posée la question ouverte sur la profession, ainsi que de différences dans la nature des variables annexes et dans la façon dont elles sont sollicitées aux deux sources. Examinons dans un premier temps la forme de la question ouverte, et les caractéristiques des réponses en clair qui lui sont données.

Question ouverte sur la profession, au recensement et à l'enquête Emploi

Recensement de mars 1982	Enquête Emploi de juin 1982
<p>(pour la profession antérieure)</p> <p>(14) AVEZ-VOUS DÉJÀ TRAVAILLÉ ?</p> <p>OUI → Quelle était votre profession principale ?</p> <p>.....</p> <p>NON</p> <p>(pour la profession actuelle)</p> <p>(15) PROFESSION EXERCÉE ACTUELLEMENT</p> <p><i>Soyez précis (Ex : électricien d'entretien de robot, comptable d'assurances technicien chimiste, etc.)</i></p> <p>.....</p> <p><i>Si vous êtes agent de la Fonction Publique, de l'État ou des collectivités (y compris HLM, hôpitaux publics), précisez votre grade (corps, catégorie...)</i></p> <p>.....</p>	<p>A1 a - Profession principale. Intitulé précis de profession.</p> <p>.....</p> <p>b - Si M. est agent de l'État, d'une collectivité locale, d'un hôpital public, d'un service public (EDF, SNCF, etc.), militaire de carrière, préciser sa classification : corps grade, etc.</p> <p>.....</p> <p>c - SI NÉCESSAIRE, éléments complémentaires précisant l'intitulé de profession.</p> <p>.....</p> <p>(suivent deux questions complémentaires pour les ouvriers ou exploitants agricoles et le personnel domestique).*</p>

* L'instruction aux enquêteurs indique :

Intitulé précis, mais concis, sans mots inutiles. Si des précisions viennent spontanément, marquez-les en A.1c et non à la suite en A.1a. Cet intitulé servira pour réaliser une codification automatique de la profession. Il ne doit pas comporter d'abréviations, ni de signes tels que parenthèses, guillemets, tirets, etc., mais peut comporter des sigles connus, SNCF, EDF, PTT, ou encore PL ou VL pour les chauffeurs. Pour les deux questions A.1a et A.1b, vous devez bannir toute indication de lieu (à la ville de ...) et en général les mentions d'entreprise, sauf si elles sont vraiment importantes (RATP, SNCF, Michelin...)

et, pour le grade :

... Il ne faut pas réécrire le grade s'il a été déjà mentionné comme profession ... Les sigles usuels sont permis, PEGC, AAP...

Ainsi les deux protocoles sont largement similaires. La différence majeure est que la question de l'enquête Emploi porte sur la profession *principale*, et que cette restriction est absente au recensement. Corrélativement, on rencontre au recensement de longs libellés, décrivant des activités multiples, difficiles à reconnaître et à coder. Au contraire, à l'enquête Emploi, les enquêteurs professionnels guident les répondants, sélectionnent et standardisent l'information qui leur paraît utile, et

enregistrent au total des libellés sensiblement plus courts, en moyenne, que ceux qui figurent sur les bulletins de recensement⁵.

Cependant la liste des libellés les plus fréquents est presque la même aux deux sources. Les libellés en clair n'étant pas disponibles pour 1982, on compare, à un niveau agrégé, les professions en clair au test du recensement de 1997 (successivement pour la profession antérieure, la profession actuelle (libellé et grade, et libellé seul) et la profession principale (libellé seul) à l'enquête Emploi de 1998. La taille des échantillons est limitée à 70 000 répondants (taille de l'échantillon de l'enquête Emploi).

Ces 70 000 répondants utilisent :

- au recensement, pour déclarer leur profession antérieure, 17 700 libellés différents, dont 13 900 occurrences uniques ;
- au recensement, pour déclarer leur profession actuelle, 20 400 libellés différents (23 400 quand on ajoute le grade dans la fonction publique), dont 16 200 occurrences uniques ;
- à l'enquête Emploi, pour déclarer leur profession principale, 15 100 libellés différents, dont 10 100 occurrences uniques.

Morphologie des libellés de profession en clair au recensement et à l'enquête Emploi

	Recensement profession antérieure	Recensement profession actuelle (libellé + grade)	Recensement profession actuelle (libellé seul)	Enquête Emploi profession principale (libellé seul)
Nombre de libellés différents	17700	23400	20400	15100
<i>fréquence moyenne</i>	<i>4,0</i>	<i>3,0</i>	<i>3,4</i>	<i>4,6</i>
Hapax (occurrences uniques)	13900	19100	16200	10100
Occurrences multiples :				
nombre de libellés	3800	4300	4200	5000
nombre de répondants	56100	50900	53800	59900
<i>fréquence moyenne</i>	<i>14,6</i>	<i>11,8</i>	<i>12,8</i>	<i>12,1</i>

Source : test recensement 1997, enquête Emploi 1998

Quand un retraité ou un chômeur déclare son ancienne profession principale, le recul qu'il a par rapport à sa situation d'emploi le conduit à fournir une déclaration plus concise, plus simple, plus vague peut-être, que s'il était au travail, d'où un taux

⁵ Si l'enquête Emploi laisse une ligne à l'enquêteur pour décrire la profession, le recensement en propose deux au recensé.

moins d'occurrences uniques et des fréquences plus élevées pour les libellés courants.

Les résultats pour l'enquête Emploi révèlent bien l'ampleur de la normalisation des libellés opérée par les enquêteurs : les hapax (libellés n'apparaissant qu'une fois) sont plus rares, mais en moyenne les descriptions sont plus concises qu'au recensement.

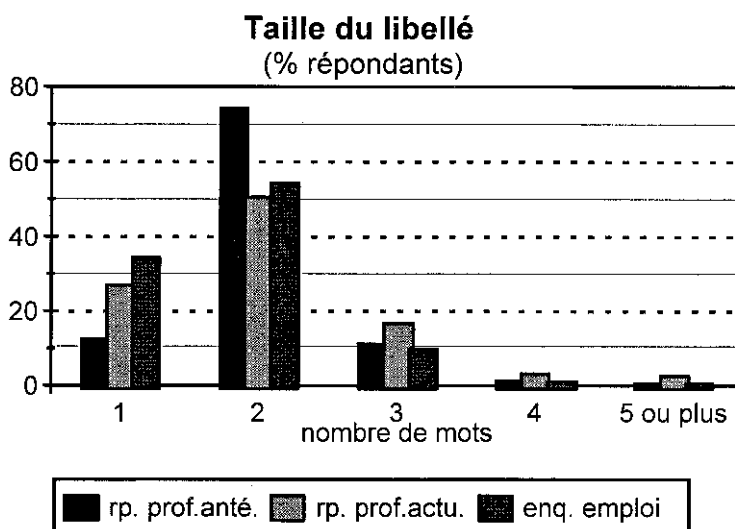
Longueur des libellés au recensement et à l'enquête Emploi (pour 1000 répondants)

Nombre de mots	Recensement profession antérieure	Recensement profession actuelle (libellé + grade)	Recensement profession actuelle (libellé seul)	Enquête Emploi profession principale (libellé seul)
1	126	229	270	343
2	743	484	505	542
3	111	179	167	98
4	13	71	32	11
5 ou plus	7	37	26	6
Total	1000	1000	1000	1000
<i>Longueur moyenne (en nombre de mots)</i>	<i>2,0</i>	<i>2,2</i>	<i>2,1</i>	<i>1,8</i>
- Hapax	2,6	3,2	2,9	2,5
- Occurrences multiples	1,9	1,9	1,8	1,7

Source : test recensement 1997, enquête Emploi 1998

La taille modale est toujours de deux. Mais les professions déclarées en deux mots représentent les trois quarts des déclarations de professions antérieures, et seulement la moitié des professions actuelles. C'est là que l'écart entre la taille des libellés rares et celle des libellés courants est la plus faible. On peut voir là aussi l'effet du temps sur la « stabilisation » de la déclaration : plus le temps passe, plus la déclaration se fige autour d'un libellé « décanté ».

Les distributions des libellés de profession actuelle au recensement sont les plus étalées : il y a deux fois plus de libellés d'au moins trois mots qu'à l'enquête emploi. Les réponses fournies par les personnes recensées sont moins standardisées que celles saisies par les enquêteurs, et comportent une plus grande part de redondance avec les informations provenant des questions ultérieures.



Neuf des dix libellés les plus fréquents sont communs aux deux listes. Parmi les 50 libellés fréquemment utilisés respectivement au recensement et à l'enquête emploi, on en retrouve 41 dans les deux sources. Ces 50 libellés sont ceux déclarés par 30 % des répondants à l'enquête Emploi, mais par 18 % au recensement. Ainsi les « secrétaires » (sans autre indication) représentent 2,6 % des répondants à l'enquête Emploi et 1,7 % au recensement.

Les 10 libellés les plus fréquents

Recensement (profession actuelle)					Recensement (profession actuelle)				
libellé	N	/1000	Rang	EE	libellé	N	/1000	Rang	RP
secrétaire	1179	16,8	1	1	secrétaire	1789	25,6	1	1
assistante maternelle	746	10,7	2	4	comptable	860	12,3	2	4
femme de ménage	573	8,2	3	3	femme de ménage	845	12,1	3	3
comptable	543	7,8	4	2	assistante maternelle	833	11,9	4	2
aide soignante	488	7,0	5	7	infirmière	769	11,0	5	6
infirmière	371	5,3	6	5	agriculteur	758	10,8	6	44
agent d'entretien	350	5,0	7	8	aide soignante	694	9,9	7	5
employée de bureau	323	4,6	8	16	agent d'entretien	687	9,8	8	7
vendeuse	323	4,6	9	10	maçon	632	9,0	9	10
maçon	315	4,5	10	9	vendeuse	617	8,8	10	9

Source : test recensement 1997, enquête Emploi 1998

Au-delà de la similitude générale, les écarts entre les effectifs observés dans les deux sources pour chaque libellé sont très révélateurs de l'incidence du protocole de collecte. On dénombre à peu près autant d'« assistantes maternelles » ou de « coiffeuses ». Ce sont des libellés qui suffisent au codage de la PCS. En revanche, on rencontre deux fois plus de « maçons », de « vendeuses », de « commerçants » ou de « menuisiers » à l'enquête Emploi. Dans la plupart des cas, le recensé indique son secteur d'activité en renseignant sa profession, alors qu'à l'enquête Emploi, l'enquêteur sait que cette information sera demandée plus loin. Au recensement l'électricien est expressément invité par l'exemple à ne pas écrire simplement « électricien », et ce libellé est trois fois plus rare à cette source qu'à l'enquête Emploi. Quant aux agriculteurs, là où le recensé précise souvent sa spécialisation, l'enquêteur note simplement « agriculteur », puis pose ensuite les questions complémentaires.

Il est donc nécessaire, pour aller plus avant dans l'explication de la variabilité des formes de déclaration et des résultats du codage, d'examiner le rôle des variables connexes. Deux questions principales permettent le codage de la catégorie socioprofessionnelle à deux chiffres, l'une a trait au *statut* (non salarié ou salarié, salarié de l'État, d'une entreprise, d'un particulier...), l'autre, posée aux seuls salariés, porte sur la *position dans l'emploi* (ouvrier non qualifié / ouvrier qualifié / employé / technicien / agent de maîtrise / ingénieur ou cadre). D'autres questions entrent surtout en jeu dans le chiffrage de la profession détaillée. On s'intéressera ici à la position dans l'emploi, dont le rôle est tout à fait stratégique. Les nombreux réaménagements qu'ont subis les questions correspondantes témoignent de difficultés qui méritent une attention particulière. Sans l'information sur la classification dans l'emploi, 13,5% des salariés au test du recensement

de 1997 - environ un bulletin sur sept - ne seraient pas codés dans la PCS qui leur est affectée sur la base de l'ensemble des renseignements collectés.

3. La position dans l'emploi, ou le statisticien embarrassé

Au recensement de 1954, la question ouverte sur la « profession principale au moment du recensement » était surmontée d'un long paragraphe de consignes invitant notamment les ouvriers à indiquer leur niveau de qualification, les fonctionnaires leur grade, etc. Parmi les personnes concernées, beaucoup omirent de donner ces renseignements ; la CSP fut chiffrée par imputation, tous les soudeurs salariés, par exemple, étant considérés comme des ouvriers qualifiés.

À partir de 1962, diverses questions fermées font suite à la question ouverte sur la profession. Une ou deux questions permettent d'abord de renseigner le statut ; en 1962, 1968 et 1975, on demande ensuite aux salariés de répondre à l'une des trois questions suivantes :

a - Si vous êtes ouvrier, précisez la qualification de votre emploi actuel (manoeuvre /OS /OQ-OHQ)

b - Si vous êtes agent de l'État, d'une collectivité locale ou d'un service public (EDF, SNCF, etc.) ou militaire de carrière, précisez votre grade (question ouverte)

c - Si vous êtes dans un autre cas, précisez votre position hiérarchique (question ouverte).

En 1975, une question supplémentaire permet de ventiler les non-salariés en fonction du nombre de personnes qu'ils emploient. En 1982, elle est intégrée dans la question sur le statut.

En 1982, la question ouverte sur la profession cesse de faire référence à la profession principale. Une même question a est posée à tous les salariés : « Indiquez la catégorie professionnelle de votre emploi actuel : ouvrier /employé / technicien, dessinateur / agent de maîtrise / ingénieur ou cadre/ autre cas ». Des ramifications répartissent les ouvriers en trois niveaux de qualification (manoeuvre /OS / OQ) et les agents de maîtrise en deux (ceux qui encadrent des ouvriers ou des employés, et ceux qui dirigent des agents de maîtrise ou des techniciens). La question b, sur le grade des agents de l'État, est inchangée depuis 1962 (mais en 1982 les agents de l'État doivent aussi renseigner la question a). Une nouvelle question fermée, sur la *fonction*, est posée aux ingénieurs, cadres, agents de maîtrise. Elle intervient dans le codage de la profession à quatre chiffres, mais pas dans celui de la catégorie à deux chiffres.

En 1990, la question sur la position dans l'emploi connaît des modifications substantielles et comporte des listes d'une longueur accrue : « Indiquez la *position* professionnelle de votre emploi actuel » : manoeuvre ou OS / OQ-OHQ / agent de maîtrise dirigeant des ouvriers... / agent de maîtrise dirigeant des techniciens / technicien, dessinateur, VRP... / instituteur, assistant(e) social(e) ... / ingénieur ou cadre / professeur et personnel de catégorie A de la fonction publique / employé de bureau ... / autre cas. La question sur la fonction propose à *tous* les salariés une liste modifiée. Une personne qui a répondu « employé de commerce » à la question ouverte sur la profession coche « employé de commerce » à la question sur la position professionnelle et « commerce » à la question sur la fonction. Le taux de non réponses est relativement élevé.

En 1999, on retourne à la formulation de 1982 : « Indiquez la *catégorie* professionnelle de votre emploi ». La liste est de nouveau modifiée : les deux premiers postes (ouvriers) sont inchangés, viennent ensuite « agent de service, aide soignant(e), employé de maison », « employé de commerce, employé de bureau, personnel administratif de catégorie C ou D de la fonction publique ». La rubrique « autre cas » disparaît, laissant dans l'embarras les policiers ou les artistes. La liste des fonctions ne subit que des retouches mineures. La redondance entre les diverses questions n'est pas diminuée.

À des variantes mineures près, les enquêtes Emploi reprennent les formulations du recensement. Tout récemment, l'enquête Emploi en continu propose une liste de positions dans l'emploi après la question « M... est-il ...? », sans donner de nom - ni « catégorie », ni « position », ni « grade », ni « classification »... - à l'ensemble des modalités.

Des tests en cours pour le recensement rénové ou pour l'enquête Emploi marquent un éventuel retour au type de questionnement des années soixante, différencié selon que le répondant est ou non salarié de l'État. Ils portent aussi sur différents ordres des modalités : si la modalité « employé » vient tôt dans une liste, elle peut être comprise comme « salarié » ; une modalité « direction générale » est moins souvent choisie si elle vient après « cadre » que si elle vient avant, etc.

L'évolution d'ensemble, notamment depuis 1982, peut être caractérisée par un allongement des listes de positions dans l'emploi, et une plus grande redondance entre les questions. Un positionnement en termes de grilles des conventions collectives était suggéré par la liste ouvrier / employé / technicien / agent de maîtrise / ingénieur ou cadre proposée en 1982. Les salariés de l'État n'y retrouvaient pas leur compte, mais ils pouvaient cocher « autre cas » et indiquer leur grade à la question suivante. Les listes ultérieures, plus copieuses et plus hétérogènes, comportent des termes qui peuvent être interprétés comme des noms de professions - voire, la culture statistique se diffusant dans le monde profane, comme des catégories socioprofessionnelles - plutôt qu'à des positions dans l'emploi ou à des classifications professionnelles proprement dites [Fermanian et Lantin, 1998]. Mais il est vrai que la seule référence aux grands postes constitutifs des grilles Parodi-

Croizat, vieilles de plus d'un demi siècle, pose problème dans un contexte où le langage des conventions de branche et d'entreprise évolue et où l'emploi de termes tels que « ouvrier », « employé », « agent de maîtrise » dans les libellés de profession est en recul [Burnod et Chenu, à paraître].

La difficulté à identifier la nature de l'information recueillie à la question sur la position dans l'emploi est patente dans le cas des techniciens. Un « technicien d'atelier » au sens de la convention collective de la métallurgie est en principe classé dans la catégorie socioprofessionnelle des ouvriers qualifiés de type industriel, et non dans celle des techniciens. Un technicien au sens des grilles d'emplois de la Sécurité sociale est un employé administratif d'entreprise au sens des CS. Une personne dont le libellé de profession est « technicien en organisation du travail » est codée dans une catégorie socioprofessionnelle de technicien, employé ou ouvrier selon la position dans l'emploi qu'elle déclare, et en employé administratif par défaut. Un « technicien en électroménager », quelle que soit la classification professionnelle qu'il déclare, est toujours codé en ouvrier qualifié. Et la formule « technicien de surface » est un libellé de profession souvent donné comme exemple d'école - en fait très rare - et conduisant à un codage non dans la catégorie socioprofessionnelle de technicien, mais dans celle d'ouvrier non qualifié de type artisanal.

La multiplication de cas témoignant de la polysémie du mot « technicien » ne met pas nécessairement en question la pertinence d'une rubrique « technicien » au sein d'une nomenclature socioprofessionnelle. Mais elle complique le travail du statisticien. À la question sur leur position professionnelle, certaines personnes peuvent répondre en des termes de jargon d'entreprise (le technicien de la sécurité sociale qui se déclare technicien), d'autre peuvent traduire ce jargon dans un langage interprofessionnel (le technicien de la sécurité sociale qui se déclare employé). Dans beaucoup de cas, les réponses aux autres questions (profession en clair, secteur d'activité...) assurent un codage socioprofessionnel satisfaisant. Mais l'ambiguïté est parfois irréductible.

Position professionnelle déclarée et codage socioprofessionnel

Test du recensement de 1997, codage semi-automatique. Champ : salariés

Position professionnelle déclarée	Groupe socioprofessionnel codé					Total	"Décalés"
	indépendants	cadres	prof.interm.	employés	ouvriers		(%)*
cadres	7	136	20	11	3	177	23
professions intermédiaires	1	7	225	17	3	253	11
employés	1	1	31	258	29	320	19
ouvriers	1	-	7	21	221	250	12
Ensemble	10	144	283	307	256	1000	//
Décalés (%)**	//	6	20	16	14	//	17

* Parmi les personnes ayant déclaré une position professionnelle donnée, part de celles qui sont classées par l'Insee dans une catégorie socioprofessionnelle dont l'intitulé est en décalage par rapport à la position professionnelle déclarée.

** Parmi les personnes classées par l'Insee dans une catégorie socioprofessionnelle donnée, part de celles qui se sont situées à une position professionnelle décalée par rapport à cette CS.

Dans 17 % des cas, il n'y a pas de correspondance simple entre la position professionnelle déclarée et l'appartenance à un groupe socioprofessionnel de la nomenclature. Ainsi, 20 % des « professions intermédiaires » au sens des catégories socioprofessionnelles sont des personnes qui ne se classent pas spontanément parmi les agents de maîtrise, les techniciens ou le cadre B de la fonction publique. 23 % des personnes qui se déclarent « ingénieur, cadre d'entreprise » ne sont pas classées parmi les cadres.

Certains des décalages entre position dans l'emploi et catégorie socioprofessionnelle relèvent du cas du « technicien de surface » qui déclare une position de technicien et que l'Insee range sans hésitation parmi les ouvriers non qualifiés. D'autres résultent de ce que certaines personnes interrogées ont du mal à comprendre sur quoi porte la question relative à la position professionnelle. Les variations dans la formulation de la question témoignent de ce que les statisticiens ont eux aussi du mal à cerner l'objet (« position dans l'emploi », ou « position professionnelle », ou « catégorie dans l'emploi », etc.). Ces difficultés sont fonction du caractère éminemment historique, daté et localisé, des termes dans lesquels employeurs et salariés conviennent de nommer les emplois et de déterminer les salaires. Il y a vingt ou trente ans, ces termes étaient souvent, dans la grande industrie, ceux des grilles Parodi-Croizat. Ils se sont toujours mal appliqués au monde de la boutique. La multiplication des conventions collectives faisant référence à des critères classants, celle des accords d'entreprise, et le recul du taux de syndicalisation des salariés s'accompagnent d'une diversification des appellations d'emploi. Une analyse plus approfondie des formes nouvelles de déclaration d'emploi et de leurs combinaisons avec les variables connexes est nécessaire à l'amélioration des chaînes de traitement statistique déterminant la position dans une nomenclature telle que la PCS.

4. Le libellé en clair, ressource majeure

Le libellé de profession en clair permet presque toujours de chiffrer une PCS à quatre chiffres, et, sept fois sur dix, la prise en compte des autres renseignements (position dans l'emploi, fonction, secteur d'activité...) n'entraîne aucune modification du codage. Ce libellé en clair est donc une ressource statistique majeure, dont des logiciels tels que Sicore permettent de tirer parti à assez peu de frais. La similitude des listes de libellés de profession en clair les plus fréquents au recensement et à l'enquête sur l'emploi de 1997-1998 indique que cette ressource est assez peu sensible au contexte de collecte, dépôt-retrait au recensement, questionnement en face à face à l'enquête Emploi.

Une des difficultés qui caractérise la collecte aux récents recensements de la population est liée aux redondances entre les différentes questions nécessaires au codage de la PCS, et aux variations dans l'interprétation de la question sur la position dans l'emploi. Dans des enquêtes avec questionnement en face à face assisté par ordinateur, l'usage de questions filtres devrait permettre à l'avenir de réduire la redondance dans le questionnement. Une plus grande priorité donnée dans les règles de codage au traitement du libellé de profession en clair permettrait de limiter les risques de divergence entre la signification du codage dans des dispositifs par dépôt-retrait et dans des enquêtes en face à face.

Références bibliographiques

Burnod G., Chenu A. (à paraître), Les représentations ordinaires de la division du travail : une étude fondée sur les déclarations de profession (communication à la journée « Sociologie des professions » organisée par la Société française de sociologie, Paris, 7 octobre 1999).

Chenu A. (1997), La descriptibilité statistique des professions, *Sociétés contemporaines*, n° 26.

Desrosières A., Thévenot L. (1988), *Les catégories socioprofessionnelles*, Paris, La Découverte.

Guglielmetti F. (2000), *L'appréhension de la profession et de la catégorie sociale dans une enquête par dépôt-retrait*, Insee, note n° 77/F420.

Fermanian J.-D., Lantin J. (1998), Conventions collectives et codage de la PCS, *Courrier des statistiques*, n° 85-86.

Héran F. (1997), La catégorie socioprofessionnelle : réflexions sur le codage et l'interprétation, in *Les indicateurs socio-politiques aujourd'hui*, L'Harmattan.

[Insee] (1981), *Les catégories socioprofessionnelles et leur repérage dans les enquêtes. Études méthodologiques*, Insee, « Archives et documents », n° 38.

Kramarz F. (1991), Déclarer sa profession, *Revue française de sociologie*, n° 32-1.

Merllié D. (1990), Les catégories socioprofessionnelles à l'épreuve de la réitération, *Population*, n° 6.

Rouault D. (1995), Les écarts d'estimation de la population active française au recensement de mars 1990 et à l'enquête annuelle sur l'emploi de janvier 1990 : d'où viennent les divergences ? in *Le recensement de la population 1990, études méthodologiques*, Insee.