

CODIFIER LA PCS A 4 CHIFFRES : UNE AMBITION RAISONNABLE ?

ANALYSE D'UN FLOU STATISTIQUE SUR LA RELATION FORMATION-EMPLOI

Françoise CEDO, Alberto LOPEZ

CEREQ, Département des Entrées dans la Vie Active

Introduction : Genèse d'un test

Les jeunes formés à la cuisine deviennent-ils cuisiniers ?

Les jeunes formés à la réparation automobile deviennent-ils mécaniciens ?

Les sortants d'une école d'architecture deviennent-ils architectes ?

Face à ce type de questions, rituellement posé au Céreq depuis ses origines, les enquêtes sur les cheminements professionnels des jeunes sortis de formation initiale sont supposées apporter de solides éléments de réponse. En effet, de telles enquêtes portent sur des échantillons importants et elles recueillent une information fine, à la fois sur la formation et sur les emplois successivement occupés. D'un côté, les données sur le niveau et sur la spécialité de formation des jeunes sont repérées à la source. L'information provient des établissements de formation d'origine, ce qui permet d'aboutir à des codes très détaillés. De l'autre côté, les jeunes sont abondamment interrogés sur leurs séquences d'emploi et les professions successivement exercées sont codées dans la nomenclature officielle la plus fine (PCS en 455 postes). Ces enquêtes semblent donc livrer « clé en main » tous les éléments nécessaires à mesurer le degré de correspondance entre formations et emplois.

Dans la pratique, une telle analyse de ces « correspondances » est excessivement complexe et délicate¹. En général, la mise en regard de données détaillées sur les diplômes et les emplois conduit souvent à de sérieux doutes sur la qualité des chaînes de traitement de formation (donc sur la fiabilité des données auxquelles on aboutit) et à des interrogations sur la pertinence même des nomenclatures.

¹ Nous ne reviendrons pas dans ce papier sur la manière d'établir un pont entre une nomenclature portant sur les formations-diplômes et une autre nomenclature sur les emplois et professions. Dans le passé, une grille avait été élaborée par le Céreq pour grouper les emplois en « familles de spécialités professionnelles » qui pouvaient être mises en regard de « spécialités de formation professionnelle » (D. Barrat et alii, 1993). De manière plus pragmatique, de nombreux auteurs, travaillant sur un champ professionnel particulier, procèdent de leur côté à des regroupements ad hoc de formations et de professions. Dans la plupart des analyses, l'usage de la nomenclature PCS constitue alors un point de passage obligé.

Cela provient parfois du repérage de quelques erreurs grossières de classement. Evidemment, toute chaîne de production a ses défaillances et ses scories. Ainsi, se retrouve classé parmi les cadres comme « professionnel navigant technique de l'aviation civile », un « pilote de ligne » détenteur d'un CAP qui travaille dans une entreprise industrielle (et qui à l'évidence pilote une ligne de production et non un Airbus). Mais le poids statistique de ces « énormités » est en réalité très faible.

De manière moins criante mais aussi moins anecdotique, quelques cas de figure entraînent une vive contestation du classement : des jeunes diplômés d'un BTS et se déclarant « électroniciens » sont classés comme « ouvriers non qualifiés »². De tels cas de figure sont plus fréquents que les précédents et peuvent de ce fait perturber l'analyse statistique de la relation formation-emploi.

Mais au-delà de ces classements surprenants, l'usage de la PCS pose d'autres types de problèmes. Ainsi, le repérage du champ professionnel correspondant à des formations professionnelles préparant aux fonctions de maintenance industrielle s'avère difficile avec le seul code PCS (cf. travaux de B. Cart et M.H. Toutain, 2002). De même, pour mieux explorer la diffusion de la fonction commerciale dans les emplois, H. Eckert et D. Maillard (2002) retournent aux libellés en clair des professions sur un champ certes balisé par la nomenclature PCS mais bien plus large que l'ensemble des postes explicitement commerciaux.

À l'évidence, un rapprochement assez fin entre formations et emplois soumet donc à rude épreuve le matériau issu de la codification des professions. C'est pourquoi, au Céreq, ce codage a toujours fait l'objet d'une attention particulière aussi bien dans la collecte que dans les traitements ultérieurs. Mais compte tenu de certaines insatisfactions des utilisateurs des données, nous avons voulu aller plus loin à l'occasion de l'enquête Génération 98³ : pour contrôler et améliorer la qualité, un test a été effectué avec une double chaîne de codification du même corpus de données correspondant à 10 000 emplois.

Ce papier est basé sur l'analyse des divergences apparues au cours de ce test. Il comprend deux parties.

La première tente d'apprécier l'ampleur et l'incidence des écarts observés entre les résultats des deux procédés de codification. Elle précise le protocole du test et le situe par rapport à d'autres expériences du même type. On cherche à voir jusqu'où le mode de codification retenu influe sur une analyse statistique des correspondances entre diplômes et débuts de carrière professionnelle. Cette partie répond à une inquiétude légitime d'utilisateur de l'enquête qui peut se demander si avec un autre procédé de codification des emplois, il aurait obtenu le même résultat.

La deuxième partie de ce papier se propose de disséquer le « flou de codification » révélé par le test. L'analyse sera d'abord menée d'un point de vue technique, en repérant les dysfonctionnements de la chaîne qui va de la confection du questionnaire au classement en PCS, en passant par l'interrogation des jeunes. Mais comme l'ont déjà révélé d'autres études de ce type (Chenu - Guglielmetti 2000, Guglielmetti 1999, Chenu 1997), les divergences et les zones d'incertitude ne sont pas distribuées au hasard dans l'espace socio-économique. C'est pourquoi, les divergences seront ensuite analysées comme un phénomène susceptible de rendre compte de la structuration plus ou moins forte des divers champs sociaux ou économiques mais aussi du « jeu » entre certains découpages fins de la PCS et les « catégories spontanées ». La spécificité du champ de l'enquête (centrée sur les jeunes en début de vie active) ne doit évidemment pas être oubliée.

² Un autre exemple concerne une « pharmacienne » n'ayant pas de diplôme supérieur au bac et classée comme « cadre » avec les « pharmaciens salariés ». Il est déjà cité dans l'ouvrage « les catégories socio-professionnelles » (Desrosières-Thévenot, 1988) pour montrer le rôle des représentations dans la chaîne de codification.

³ Génération 98 est le nom de l'enquête sur les cheminements professionnels des jeunes sortis de formation initiale en 1998. Il ne s'agit pas d'une enquête auprès d'une génération d'âge mais d'une promotion de sortants. C'est la deuxième enquête du Céreq portant simultanément sur l'ensemble des niveaux de formation initiale.

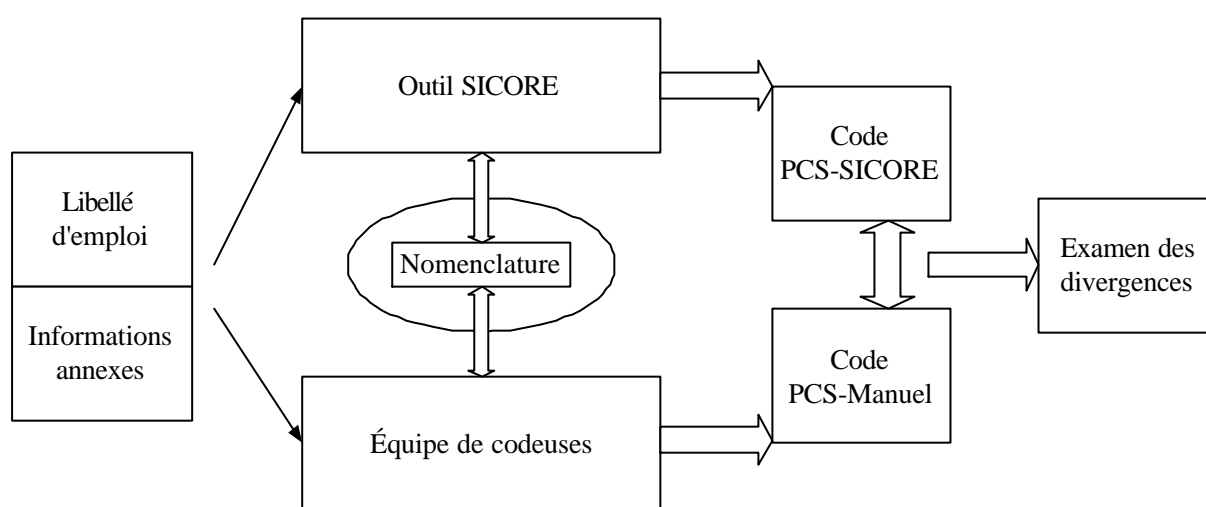
1. L'épreuve d'un test de qualité

1.1. Une expérience basée sur la confrontation entre l'homme et la machine

Pour l'enquête Génération 98, comme pour la plupart des enquêtes du Céreq ou de l'INSEE, la codification de la PCS articulait deux modes de codification : un mode automatique avec l'outil SICORE (Système Informatique de CODage par REcognition) et un mode dit « manuel »⁴, reposant sur une petite équipe de « codeuses » préalablement formées⁵.

Dans l'organisation habituelle, cette équipe intervient seulement sur les emplois que la machine n'a pas réussis à codifier (on parle de « reprise manuelle »). Mais pour contrôler la qualité de la PCS, un échantillon de 10 000 emplois préalablement codifiés par SICORE, a fait l'objet d'une codification manuelle⁶. (graphique 1)

Graphique 1 : Protocole du test



Avant de préciser la manière dont a fonctionné chacun des deux procédés, il est utile d'explicitier ce qui constitue la matière première des deux classements.

1.1.1. Un même matériau pour le classement

La codification en PCS nécessite de recueillir non seulement un intitulé de profession très précis mais aussi bien d'autres informations. Parmi celles-ci, certaines sont destinées à préciser la position de l'individu : le statut (salarié / indépendant), la classification professionnelle dans les conventions collectives, la fonction principale exercée. D'autres informations annexes portent sur l'employeur précisant sa nature (distinction public / privé), son activité économique et la taille de l'entreprise. A l'enquête Génération 98, pour chaque emploi occupé par les jeunes interrogés, on a cherché à recueillir cet ensemble d'informations avec un questionnaire assez proche des principales enquêtes statistiques auprès des ménages (voir en encadré le détail).

⁴ Nous reprendrons ici ce terme de « manuel » peu adapté à une opération qui comme nous le verrons mobilise peu les « mains » et beaucoup « l'intellect », les systèmes de « représentation cognitive » (comme le montrent bien Desrosières et Thévenot), la culture générale, etc.

⁵ Au Céreq, la codification manuelle est sous-traitée à une société extérieure qui travaille sur la base d'un cahier des charges en bénéficiant d'une formation et d'un accompagnement.

⁶ Des tests de même type, permettant une confrontation entre les résultats d'une double codification des emplois des mêmes individus, ont déjà été effectués dans le passé sur le recensement de la population ou/et l'enquête emploi. Ils seront évoqués dans le papier lorsque des comparaisons semblent possibles.

Le recueil d'informations utiles à la PCS dans l'enquête Génération 98

Dans l'enquête téléphonique Génération 98, un module de questions était appliqué pour chaque emploi occupé par un individu depuis la sortie du système éducatif jusqu'à la date d'interrogation (avril-juillet 2001). Le questionnement et les premiers traitements ont été conçus pour permettre un chiffrage de la profession avec le logiciel SICORE. La partie du questionnaire correspondant à cette finalité est voisine de celle des Enquêtes Emploi de l'INSEE mais quelques écarts par rapport aux procédures habituelles de l'INSEE méritent toutefois d'être signalés.

D'une part, dans le questionnaire de Génération 98, la classification professionnelle et la fonction principale sont ainsi un peu plus détaillées, pour faciliter un éventuel codage manuel ultérieur.

D'autre part, dans Génération 98, le secteur d'activité de l'entreprise n'est pas codé aussi finement et fiablement, en NAF, que dans l'enquête emploi. Dans cette dernière, on recueille les noms et adresse de l'employeur, on passe par un appariement avec le répertoire des entreprises (SIRENE) et on attribue ainsi une activité économique très détaillée à chaque emploi. L'enquête Génération 98, de son côté ne recueille pas l'adresse de l'employeur à chaque emploi, mais, avec un autre procédé, elle aboutit à une codification de l'activité économique de l'employeur en 65 postes, en NES. Une imputation du code NAF a été opérée à partir de la NES pour permettre à SICORE de « tourner »⁷. Comme nous le verrons, cela est évidemment susceptible de générer un aléa dans la codification automatique sur certains champs.

Dans le test qui a été organisé, toute cette « matière première » du classement en PCS a été livrée selon des modes évidemment différents d'un côté et de l'autre. Alors qu'un fichier de codes a été livré à la machine, l'équipe de codification manuelle a reçu pour chaque emploi à traiter, une fiche papier sur laquelle figuraient, outre l'information codifiée nécessaire à SICORE, des informations en clair comme le nom de l'employeur, l'activité économique de celui-ci telle qu'elle avait été formulée par l'enquêté, etc. Dans ce duel homme-machine (ou plutôt femme-machine), la machine semble légèrement désavantagée au départ, mais précisons bien qu'en aucun cas, les codeuses ne disposaient des variables sur le diplôme du titulaire du poste et son salaire (variables qui apporteraient de précieux indices mais qui n'ont pas été retenues par les concepteurs de la nomenclature). Pour un emploi donné, elles n'avaient pas connaissance, non plus, des autres emplois occupés par le même jeune, antérieurement ou postérieurement.

Au final, le matériau fourni à l'outil de codification automatique et aux codeuses peut donc être considéré comme très voisin.

1.1.2. Une référence commune pour les règles de codification

De même, les deux modes de codage (automatique et manuelle) ont une référence commune inscrite dans un volumineux document intitulé « Nomenclature des professions et catégories socioprofessionnelles, 1994, PCS » qui livre les instructions de bases ainsi qu'un « index analytique » précisant pour chaque PCS le « noyau », les « assimilés » et les « cas limite exclus ». Mais à partir de cette même référence, l'homme et la machine vont fonctionner avec des ressources et des logiques sensiblement différentes.

⁷ On a simplement affecté chaque poste NES sur le poste NAF le plus courant au sein de celui-ci.

1.1.3. ... Mais des fonctionnements différents

La codification automatique de la PCS repose sur une transcription des instructions en un ensemble de « règles dures ». Plus que l'humain, le programme s'appuie avant tout sur la reconnaissance d'un libellé de profession⁸ et la prise en compte systématique du statut. Ainsi, SICORE n'a mobilisé la « position professionnelle » que dans un cas sur deux, la nature de l'employeur (distinction public / privé) dans un cas sur quatre et le secteur d'activité dans un cas sur dix.

De leur côté, les codeuses tiennent davantage compte de l'ensemble des variables disponibles. Certes, elles commencent par une lecture du libellé et disposent des règles de chiffrement privilégiant une entrée alphabétique de libellés. Mais à la différence du logiciel SICORE, cet « index alphabétique » n'a pas été actualisé et en 2002 on utilise toujours l'Édition de 1983 ! Les indices d'obsolescence de ce document ne manquent pas : on y trouve par exemple, pour de nombreux cas de figure, une correspondance entre rubriques PCS et postes de la NAP, nomenclature d'activité ayant été remplacée par la NAF. Les codeuses doivent donc davantage se référer à l'index analytique (source) et faire preuve de synthèse.

Autre différence dans le fonctionnement des deux modes de codage :

D'une certaine manière, SICORE capitalise les expériences de chiffrement précédentes, dans la mesure où l'INSEE en transcrit les enseignements en modifiant l'outil. Plusieurs versions de SICORE se sont succédées. Les bases de connaissance ont été enrichies par de nouveaux libellés et le système de « règles dures » a également évolué au cours du temps. À défaut d'une mise à jour de l'index alphabétique de 1983, du côté des codeuses, seules quelques consignes (dispensées en formation ou cours de traitement), viennent constituer une jurisprudence sur des cas problématiques revenant assez fréquemment.

On perçoit donc comment, avec un matériau presque identique et avec une même référence pour les règles de codage, l'homme et la machine peuvent arriver à des résultats différents.

1.2. L'ampleur des divergences

1.2.1. Une discordance sur 35% des emplois

Sur les 10 000 emplois testés, les divergences entre le code automatique et le code manuel sont assez fréquentes. Pour 35 % des emplois, les deux codes PCS fins (qu'on appellera « codes profession ») ne concordent pas. Au fur et à mesure qu'on agrège la nomenclature, ce taux de divergence⁹ diminue. Il est de 19 % sur la CS à 42 postes (qu'on appellera « catégorie socioprofessionnelle ») mais reste encore à 10 % sur la CS à 8 postes (qu'on appellera « groupe social »).

L'ampleur de ces divergences peut surprendre et inquiéter les utilisateurs de l'enquête. Mais comparons nos résultats avec ceux obtenus par l'INSEE, dans des opérations analogues.

1.2.2. Un résultat comparable à celui d'autres expériences

En se référant à la communication d'A. CHENU et F. GUGLIELMETTI présentée aux journées de méthodologie statistique de décembre 2000¹⁰, le taux de 35 % de divergences sur la PCS dans notre

⁸ Cette « reconnaissance », basée sur une analyse en bigrammes des libellés, est plus ou moins performante avec un risque de « confusion » pour des énoncés d'emploi assez longs.

⁹ Les taux de divergences sont ici calculés avec une pondération qui permet de tenir compte du taux de sondage inégal qui a été pratiqué sur l'échantillon initial et de la pondération initiale appliquée aux répondants.

¹⁰ D'autres chiffres concernant les tests du RP99 ont été publiés et semblent contradictoires avec ceux-ci. Mais nous avons choisi la source qui comparait plusieurs tests.

test serait intermédiaire entre le taux de 43 % de divergences au rapprochement de 1982 (entre enquête emploi et Recensement) et le taux de 29 % au rapprochement de 1997 (entre codifications automatique et manuelle sur le test du recensement).

Le premier taux (1982) combine l'effet d'un « aléa de déclaration » avec celui du « mode de traitement de l'information » puisque l'information traitée de part et d'autre proviennent de deux déclarations différentes. Il est donc assez logiquement supérieur. Pour ce qui est du deuxième taux (1997), il porte sur un stock d'emplois occupés par une population active de tout âge. A priori, il s'agit donc d'emplois, en moyenne moins fugaces, plus stabilisés que ceux qui ont été traités dans notre test.

Ainsi, ces quelques comparaisons nous révèlent que cette « instabilité de codage », loin d'être propre à Génération 98, concerne dans des proportions équivalentes, les autres grandes enquêtes du système statistique public.

Cela peut certes rassurer les producteurs des enquêtes Génération mais l'utilisateur des données soucieux d'analyser les liens formation-emploi reste, à ce stade, assez dubitatif. Aboutira-t-il à des conclusions différentes suivant que les emplois auront été codifiés par SICORE ou de façon manuelle ? Et si oui, quelles conclusions pourront être considérées comme valides ?

Pour répondre à ces questions, nous nous sommes livrés à quelques simulations.

1.3. Sensibilité de la mesure des liens formation-emploi à la méthode de codification

Disposant des deux jeux de données issus des deux codifications sur les 10 000 emplois traités, il est facile de se livrer à un diagnostic comparé sur l'ampleur des « déclassements » mesuré d'un côté et de l'autre. Pour ce faire, nous nous plaçons dans le cadre d'une simple analyse de la position professionnelle des emplois par grands niveaux de formations.

1.3.1. Une fréquence des déclassements peu tributaire du mode de codification

Les écarts observés sur les indicateurs choisis sont parfois significatifs mais ils restent faibles. Par exemple, pour les bac+2, malgré de fréquentes divergences entre les deux codifications, la part des emplois de niveau « employé-ouvrier » est de 35 % à l'issue de SICORE, contre 33 % à l'issue du codage manuel. De manière générale, le codage manuel tend à réduire la fréquence des déclassements pour les diplômés du supérieur et des sur-classements pour les moins diplômés. Mais tous ces écarts sont en général inférieurs à l'intervalle de confiance lié au sondage. (tableau 1)

Tableau 1 : Part des employés/ouvriers chez les bacheliers¹¹

	Part SICORE (en %)	Part manuelle (en %)	Ecart en points	Significativité du test
bac pro et techno	81,6	80,7	-0,9	NS
bac + 1, bac + 2 non diplômés	63,8	61,6	-2,2	NS
bac + 2	35,1	32,8	-2,3	--
2nd cycle	23,5	20,7	-2,8	--
3e cycles et écoles	5,3	4,2	-1,1	--

Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

¹¹ Taux calculés sur 10.000 emplois codés deux fois, par SICORE et en manuel, en PCS

1.3.2. Impact limité sur les correspondances entre spécialités de formation et profession exercées

L'effet du type de codage peut également être testé sur une analyse des correspondances entre spécialités de formation et domaine professionnel des emplois. (tableau 2)

Tableau 2 : Destination professionnelle des diplômés selon la méthode de codification PCS - Principaux domaines professionnels pour quelques spécialités de formation¹² (1)

Spécialité de formation	Domaine professionnel	Part SICORE (%)	Part manuelle (%)	Ecart en points	Significativité du test
<i>Moteurs et mécanique auto</i>	13- Métallurgie, mécanique	55,2	56,4	1,2	NS
	32- Transport, logistique	16,2	14,9	-1,3	NS
	20- Bâtiment, aménagement	6,6	5	-1,6	NS
	14- Energie, système électrique	5,4	4,1	-1,3	NS
<i>Electricité, électronique</i>	14- Energie, système électrique	32,5	38,8	6,3	++
	13- Métallurgie, mécanique	12,0	11,8	-0,2	NS
	32- Transport, logistique	11,8	10,8	-1,0	NS
	16- Informatique	8,7	9,2	0,5	NS
	99- Domaine NC	7,7	4,9	-2,8	--
	20- Bâtiment, aménagement	6,1	3,8	-2,3	--
<i>Agro-alimentaire, alimentation, cuisine</i>	11- Ind. agricole et alimentaire	63,0	60,4	-2,6	NS
	32- Transport, logistique	10,1	9,1	-1	NS
	39- Hébergement, tourisme	5,3	10,7	5,4	++
<i>Comptabilité-gestion</i>	31- Gestion, comptabilité	19,0	19,0	0,0	NS
	99- Domaine NC	14,2	13,7	-0,5	NS
	32- Transport, logistique	14,0	15,3	1,3	NS
	33- Commerce	11,9	12,1	0,2	NS
	35- Info-communication	11,1	11,6	0,5	NS
	39- Hébergement, tourisme	7,9	7,1	-0,8	NS

"Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

Là encore, il apparaît que les estimations et le diagnostic varient assez peu en fonction du procédé de codification retenu. Par exemple, les jeunes formés dans le champ « comptabilité - gestion » ont leurs emplois à 19 % dans le domaine professionnel « gestion - finances - comptabilité » aussi bien par SICORE que par la codification manuelle.

Souvent, aucune différence significative n'apparaît entre les deux résultats. Les principaux domaines d'emploi sont toujours les mêmes avec une part équivalente. Toutefois, le cas de quelques spécialités de formation mérite d'être cité. Par exemple, les emplois des jeunes sortant des formations du domaine « électricité, électronique » se trouvent davantage concentrés dans le premier domaine professionnel correspondant (« énergie et systèmes électriques ») par la codification manuelle (39%) que par SICORE (33 %). Il en est de même pour les formations à l'hôtellerie et au tourisme (50 % contre 44 %), à l'informatique (63% contre 59 %) et au secrétariat (31% contre 27 %).

¹² Taux calculés sur 10.000 emplois codés deux fois, par SICORE et en manuel, en PCS

Là où des écarts significatifs apparaissent, la codification manuelle aboutit le plus souvent à un lien un peu plus serré entre formations initiales et emplois¹³.

A l'issue de cette simulation, l'utilisateur des données de Génération 98 peut donc être plutôt rassuré. Pour peu que la PCS soit utilisée à des fins statistiques sur des populations numériquement assez importantes, les indicateurs usuellement calculés pour analyser les liens formation-emploi ne varient pas fortement en fonction du mode de codage. Les variations sont souvent d'une ampleur comparable à l'imprécision liée au sondage.

La situation ne peut être jugée comme satisfaisante pour autant. Même si les divergences à l'issue des deux modes de codification semblent se compenser statistiquement, il reste qu'elles introduisent l'idée qu'il y a un aléa plus ou moins fort dans la codification. Mais d'où provient exactement ce flou ? Peut-il être considéré comme le résultat d'un simple dysfonctionnement d'une chaîne de traitement de l'information ? Ne faut-il pas essayer de tenir compte du jeu subtil à travers lequel se négocient des positions et des identités professionnelles ?

2. Anatomie d'un flou

Le « flou statistique » mérite d'être analysé sous divers angles.

2.1. Le diagnostic du technicien

Le premier diagnostic sollicité est celui du « technicien ». Quels sont les maillons défaillants dans cette chaîne qui va de la conception du questionnaire au code PCS ? Et tout d'abord, ne faut-il pas remettre en cause l'homme ou la machine ?

2.1.1. Un arbitrage renvoyant dos à dos l'homme et la machine

L'expertise des divergences par une personne rompue à la nomenclature et à la codification en PCS depuis plusieurs années est évidemment précieuse. Les quelques 2000 divergences sur la catégorie socioprofessionnelle (42 postes) ont donc été examinées par un tel « expert PCS » qui a rendu un arbitrage en les classant en trois grandes catégories :

- erreurs SICORE,
- erreurs manuelles,
- incertitudes de codage¹⁴.

A l'issue de cette opération, les « erreurs SICORE » semblent un peu plus fréquentes que les « erreurs manuelles » mais surtout, les cas d'incertitude apparaissent aussi nombreux que l'ensemble des cas d'erreurs : 47 % des divergences examinées rentrent dans cette catégorie.

Ces résultats doivent être relativisés et interprétés avec beaucoup de prudence. Dans les cas où l'appellation de profession est répertoriée dans l'index alphabétique ou que des règles de décisions existent, sont claires, cohérentes et compréhensibles, l'expert peut facilement conclure à une erreur de codage, soit de la machine (SICORE), soit de l'homme (codeur manuel). Mais dans les cas où l'appellation de profession est absente des bases de connaissance de la nomenclature alors, il n'existe pas de règles dures directement applicables. L'arbitrage de l'expert devient plus délicat.

¹³ Il est important de rappeler que les codeuses ne disposaient pas d'information sur le diplôme des jeunes dans leurs fiches.

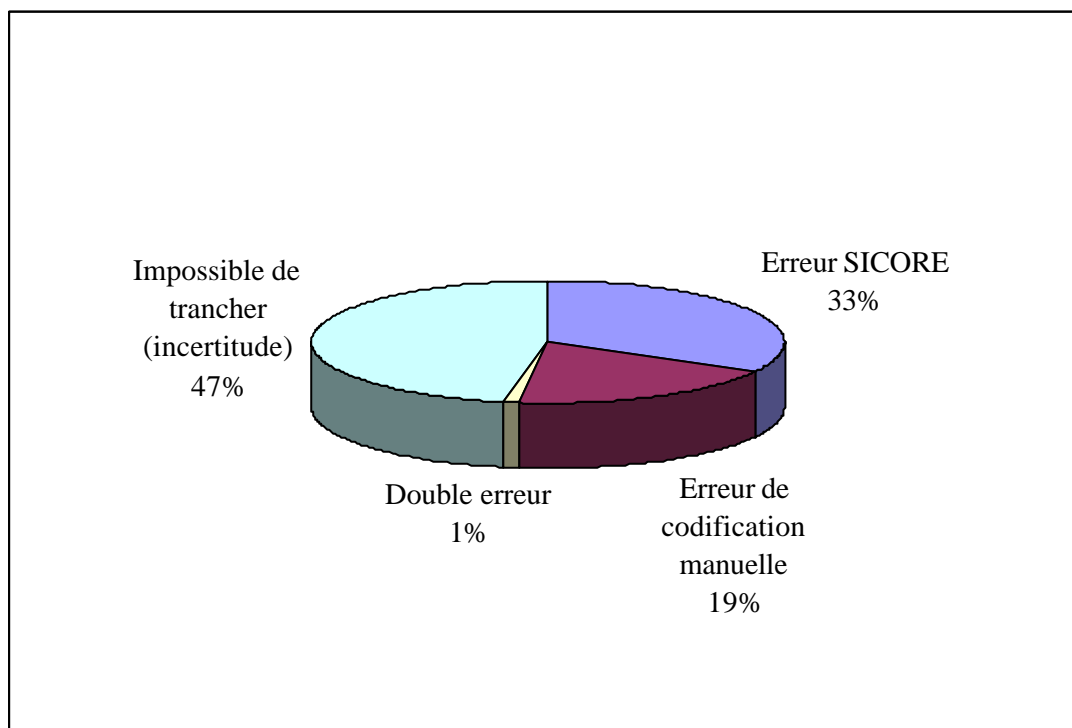
¹⁴ Les cas où les codes SICORE et manuels ont été jugés erronés sont marginaux.

Il peut mettre en jeu sa propre interprétation des règles de référence ainsi que des consignes non formalisées par l'INSEE mais retenues pour le chiffrage manuel¹⁵. Il peut ainsi favoriser les résultats issus de la codification manuelle. Par ailleurs, à partir de quel moment, jugera-t-il qu'il est impossible d'arbitrer entre les deux codes et estimera-t-il qu'il se retrouve confronté à une incertitude de codage ? La frontière entre « incertitudes » et « codes erronés » est donc elle-même très subtile. Compte tenu de tous ces risques de partialité, le verdict d'un jury (avec deux ou trois experts) aurait sans doute été plus solide que celui d'un seul expert.

2.1.2. Une masse de cas incertains ...

Il reste que dans un cas de divergence (CS42) sur deux, une « personne qualifiée » estime qu'il serait impossible de classer l'emploi dans les catégories socioprofessionnelles de façon fiable et certaine, compte tenu de l'information recueillie dans l'enquête. D'une certaine manière, cela renvoie dos à dos l'homme et la machine en incriminant à ce stade tout autant la qualité du matériau de base que les méthodes de traitement de l'information recueillie. (graphique 2)

Graphique 2 : Examen des divergences au niveau CS 42



¹⁵ D'autant plus que l'expert est souvent intervenu préalablement pour former l'équipe de codeuses.

2.1.2.1. D'abord liée au poids des libellés flous

L'analyse des libellés les plus fréquemment rencontrés dans les divergences vient corroborer et affiner ce premier diagnostic. (tableau 3)

Tableau 3 : Libellés d'emploi apparaissant plus de 20 fois dans les divergences

Libellés	Nb de divergences PCS	Taux de divergence au niveau du code profession :		
		GS ⁽¹⁾	CS ⁽²⁾	PCS ⁽³⁾
Vendeur	103	8%	9%	54%
Ouvrier	70	3%	25%	52%
Aide éducateur	69	4%	96%	96%
Electricien	66	24%	55%	59%
Assistant commercial	57	10%	12%	70%
Technicien	48	10%	26%	59%
Commercial	47	8%	20%	64%
Manœuvre	39	3%	35%	62%
Technicien de maintenance	39	11%	11%	61%
Agent administratif	37	3%	10%	62%
Employé	37	22%	60%	82%
Opérateur	36	9%	20%	82%
Technicien de laboratoire	35	2%	8%	71%
Agent de fabrication	32	3%	9%	47%
Chargé de mission	32	33%	56%	74%
Contrôleur de gestion	32	0%	12%	94%
Commis de cuisine	31	67%	67%	67%
Agent de maintenance	28	8%	51%	76%
Chargé d'études	27	0%	42%	54%
Dessinateur	27	17%	21%	64%
Mécanicien	27	1%	20%	27%
Monteur	27	0%	8%	75%
Agent de production	26	0%	8%	44%
Assistant de direction	25	60%	60%	71%
Ingénieur d'études	25	3%	16%	32%
Secrétaire	25	14%	14%	14%
Animateur	24	21%	27%	27%
Employé polyvalent	23	4%	72%	92%
Chauffeur	22	0%	43%	96%
Agent commercial	21	68%	72%	84%
Enseignant	21	23%	23%	33%
Professeur	21	11%	12%	12%
Télévendeur	21	0%	75%	88%
Rappel des taux pour l'ensemble	10 000	10%	20%	35%

(1). GS : groupes sociaux en 8 postes ou 1er chiffre du code PCS

(2). CS : catégories sociales en 42 postes ou 2 premiers chiffres du code PCS

(3). PCS : catégories socio-professionnelles à 4 chiffres en 455 postes

Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

En effet, la liste de ces libellés fait avant tout apparaître un ensemble de libellés flous, ou du moins trop imprécis au regard des exigences de la nomenclature. Dans ces cas, le classement va se jouer sur la synthèse des réponses aux questions annexes telles que le niveau de qualification, la fonction exercée, le secteur d'activité de l'entreprise, la nature de l'employeur...

En réalité, si on se positionne par rapport aux consignes de collecte et aux règles de codification, le degré de flou de ces intitulés est assez disparate.

Certains énoncés d'emploi auraient dû faire l'objet d'une relance de la part de l'enquêteur. Ainsi étaient a priori proscrites les réponses telles que « employé », « ouvrier » ou « technicien ». Mais le contrôle de qualité sur un plateau d'une centaine de télé-enquêteurs n'est pas toujours aisé. Au final, on retrouve en nombre, plusieurs autres libellés flous qui devaient être évités, à la fois du côté tertiaire (vendeur, agent administratif, commercial, assistant commercial) et du côté industriel (opérateur, manœuvre, monteur, chauffeur...). Un vendeur ou un commercial est codé en fonction du produit qu'il vend ; pour un ouvrier ou un technicien, la nature des tâches exécutées doit être précisée. En l'absence d'information suffisante sur les libellés, les deux méthodes de codage procèdent différemment. SICORE applique des règles par défaut. Le codage manuel prend davantage en compte la fonction (comptabilité, vente, entretien, fabrication ...) et l'activité économique (commerce de détail, postes et télécommunications, industrie automobile, plasturgie...). Résultat : pour beaucoup de libellés, la divergence reste fréquente jusqu'au niveau de la CS 42.

D'autres énoncés d'emploi ne faisaient pas l'objet de consignes précises aux enquêteurs. Ils peuvent paraître assez précis mais, en réalité, sont insuffisants compte tenu des exigences de la codification. Ainsi, il n'est pas facile de coder les « ingénieurs d'études » et les « chargés d'études », même avec les renseignements sur la fonction et le secteur d'activité. Pour les ventiler entre la catégorie des cadres d'entreprises (CS 37) et celle des ingénieurs et cadres techniques (CS 38), il faut une information plus fine (préciser ce qu'ils « étudient ») que seule l'énonciation de profession peut apporter. Pour certaines appellations d'ouvriers - dont la plus fréquente est « électricien » -, on peut également estimer que les informations collectées ne permettent même pas de classer de façon fiable dans une catégorie socioprofessionnelle. Les distinctions « qualifié/non qualifié » et « type artisanal / type industriel » se révèlent assez aléatoires.

2.1.2.2. Quelques incertitudes de classement liées à l'innovation

Au sein de la liste des divergences les plus fréquentes, peu de libellés semblent correspondre à de « nouveaux métiers ». Cela s'explique facilement puisque les nouveaux libellés ne sont pas souvent reconnus par SICORE (et pour cause). Ils se trouvent donc dans les « rejets de SICORE » et de ce fait, n'entrent pas dans le champ de la comparaison. Trois professions émergentes et en pleine évolution ont, malgré tout, été codées par SICORE. Il s'agit des « aides éducateurs », des « animateurs » et des « télévendeurs ».

Alors que les métiers du secteur social et de l'animation se sont beaucoup diversifiées, la nomenclature de 1982 n'identifiait qu'une seule catégorie d'aides éducateurs. Ceux-ci étaient codés avec les « éducateurs spécialisés », travaillant dans des établissements de type médico-éducatif. Mais en 1999, dans le cadre du programme « nouveaux emplois-nouveaux services » sont apparus de nombreux jeunes « aides éducateurs » exerçant dans des collèges ou des lycées publics. Ignorant cette innovation, SICORE a appliqué la règle ancienne sur ce libellé alors qu'il est évidemment nécessaire d'introduire une nouvelle règle de décision. Du côté de la codification manuelle, les nouveaux « aides éducateurs » ont été repérés, rapprochés des animateurs de l'éducation nationale et classés avec les « conseillers d'éducatifs et surveillants ». Cette option semble confirmée par la nouvelle nomenclature PCS (adoptée en 2001) qui comporte un poste « Surveillants et aides éducateurs des établissements d'enseignement ».

De même dans l'index alphabétique de l'INSEE, on ne trouve aucune règle concernant les « animateurs ». Doit-on les associer aux professions intermédiaires de l'enseignement ou à celles de la santé et du travail social? Par rapport à SICORE, le codeur manuel utilise une indication précieuse mais non codifiée : le « nom de l'entreprise ». Cette information permet alors de classer de manière plus précise les « animateurs » et plus largement les formateurs et les enseignants.

Les télévendeurs sont apparus après 1982. Pour cette nouvelle profession, l'INSEE avait ajouté dans SICORE des règles mais elles n'ont pas été portées à connaissance et officialisées. Le flou des règles de décision occasionne ici une incertitude de codage. Doit-on coder ces télévendeurs du côté des « employés administratifs » ou bien du côté des « vendeurs » ? L'arbitrage ne peut être fait sur la base du secteur d'activité, le plus souvent cité « *publicité, études de marché, sondages* ». La nouvelle nomenclature PCS est venue apporter une réponse en créant un poste « *Vendeurs par correspondance, télévendeurs* ».

2.1.2.3. Des zones d'incertitudes également liées aux règles incomplètes

Plusieurs intitulés, correspondant à des emplois moins « nouveaux » renvoient eux-aussi à des zones d'incertitude ou d'obsolescence des règles de codification.

L'« agent commercial » en est l'exemple le plus caractéristique. Même le premier chiffre est difficile à déterminer avec certitude. S'agit-il d'un cadre, d'un représentant ou d'un employé ? La majorité des jeunes agents commerciaux se sont dits « employés » et ont été codés soit avec les professions intermédiaires commerciales par SICORE, soit avec les employés administratifs (banque, assurance, transport ...) par la SAMO. Difficile de trancher dans la mesure où les règles de décision sont imprécises. On ne peut que constater l'existence d'une zone d'indécision entre deux catégories.

De même, pour les « assistants de direction », que faire lorsqu'ils (ou elles) se déclarent de niveau « employé » ? L'index alphabétique de 1983 ne livre qu'une règle, propre au secteur de la banque. Pour le reste, doit-on systématiquement les considérer comme « professions intermédiaires » ou doit-on tenir compte de la « position professionnelle » déclarée ?

De manière générale, dans les règles écrites, la prise en compte de la réponse à la question sur cette « position professionnelle » varie d'un libellé à l'autre sans qu'une logique très claire apparaisse. Le « commis de cuisine » en fait les frais. Lorsque le titulaire du poste se positionne comme « employé », SICORE applique la « règle dure » qui les classe, avec les cuisiniers, comme ouvriers qualifiés (CS 63). Le codeur manuel, jugeant cette règle contraire au sens commun, les classe avec les employés de l'hôtellerie (CS 56).

2.1.3. La faiblesse du matériau collecté

Bien que restreinte par rapport à l'ensemble des divergences¹⁶, cette analyse des cas fréquents (tableau 4) nous renforce dans l'idée que, malgré les consignes aux enquêteurs, la matière première collectée auprès des enquêtés reste en deçà de la qualité attendue. Du point de vue opérationnel, on peut évidemment en conclure que les consignes devraient encore être plus détaillées pour gagner en « précision » dans les libellés d'emploi et surtout que la surveillance du respect de ces consignes devrait être plus serrée. On peut par exemple imaginer que dans une interrogation assistée par ordinateur, le programme CATI empêche le télé-enquêteur de poursuivre son interrogation tant qu'il n'a pas ajouté une précision à certains « libellés incomplets ».

¹⁶ Au total, les libellés fréquents étudiés ici rassemblent 23 % des discordances au niveau du code profession. Le tableau 4 montre que moins un libellé est fréquent et plus il a une probabilité de divergence élevée. Le diagnostic pourrait donc être un peu affiné avec un échantillon de ces « divergences plus rares ».

Mais ces recommandations techniques, dignes des plus astucieux ingénieurs tayloriens, ne renvoient-elles pas vers d'autres obstacles plus fondamentaux liés aux difficultés d'une catégorisation ?

Tableau 4 : Taux de divergences PCS en fonction de la fréquence des libellés

Fréquence du libellé sur les 10 000 emplois	Nb de divergences PCS	Taux de divergence au niveau du code profession :		
		GS ⁽¹⁾	CS ⁽²⁾	PCS ⁽³⁾
1 ou 2	1068	15%	29%	48%
3 à 5	1166	12%	27%	44%
6 à 9	983	12%	24%	40%
10 à 20	1248	10%	19%	36%
20 à 49	2230	10%	18%	33%
50 ou plus	3305	6%	15%	27%
Ensemble	10000	9%	19%	35%

(1). GS : groupes sociaux en 8 postes ou 1er chiffre du code PCS

(2). CS : catégories sociales en 42 postes ou 2 premiers chiffres du code PCS

(3). PCS : catégories socio-professionnelles à 4 chiffres, en 455 postes

Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

2.2. L'énoncé de son emploi : un acte socialement conditionné

On conviendra aisément que la notion de « flou » et de « précision » dans l'énoncé d'un emploi est très relative. Dans la posture de technicien que nous avons adoptée jusqu'ici nous avons peu remis en cause l'objectif de l'opération et le classement dans la nomenclature PCS a été a priori considéré comme un processus technique. Mais les faits sont têtus et la matière humaine risque bien de résister à cette entreprise. C'est pourquoi il faut maintenant revenir sur le processus de classement et le considérer comme un « acte social » intervenant dans un champ plus ou moins structuré, dans lequel, même l'énoncé d'un emploi peut être soumis à de multiples tensions.

Comme l'a fort bien analysé F. Kramartz (1991), de nombreux éléments conditionnent le recueil d'un énoncé de profession. Tout d'abord, pour livrer un intitulé d'emploi, un enquêté va puiser dans un ensemble de « ressources ». Il peut mobiliser le registre technique (référence à la machine, au produit fabriqué, à la fonction, la compétence ou l'organigramme dans une organisation du travail), le registre administratif (référence aux classements issus des conventions collectives, aux statuts, aux grilles de la fonction publique). Ensuite, en livrant un libellé d'emploi, il peut aussi exprimer un attachement personnel dans ce qui est une épreuve sociale de « présentation de soi ». Enfin, on comprend donc que l'énoncé recueilli dépendra du cadre dans lequel est interrogé l'individu.

2.2.1. D'OS à opérateur : identité professionnelle et organisation du travail

Ainsi, l'analyse des différentes appellations utilisées par les jeunes ouvriers de l'automobile, pour décrire leur profession, montrent que dans bien des cas, un intitulé flou pour la nomenclature ne l'est pas pour l'enquêté. Sur les dix énoncés de profession les plus fréquents, sept sont flous pour la PCS¹⁷. Mais les 20 % de déclarations de type « agent de fabrication »¹⁸ font implicitement référence aux conventions collectives de branches. Les déclarations de type « ouvrier spécialisé » ou « ouvrier(e) » (7%) peuvent s'interpréter comme l'affirmation d'appartenance à un groupe social. Enfin les déclarations plus rares de type « opérateur » (3%) correspondent à un effort de positionnement dans

¹⁷ Les sept intitulés flous sont : « agent de fabrication », « agent de production », « ouvrier(e) », « opérateur », « ouvrier spécialisé », « ouvrier à la chaîne », « ouvrier de fabrication ».

¹⁸ Type « agent de fabrication » : « agent de fabrication », « agent de production », « ouvrier de production », « ouvrier à la chaîne ».

l'organisation du travail. Celle-ci s'est considérablement transformée à partir des années 1980, dans les usines Peugeot notamment, avec l'automatisation et les systèmes de production en continu. Dans le travail de montage, la phase de contrôle est de plus en plus intégrée à la phase de production.¹⁹ Les employeurs cherchent à valoriser le poste de travail : pour désigner un « OS de chaîne », le terme d'opérateur apparaît. Certaines conventions collectives ont ainsi modifié leur grille, en particulier en ce qui concerne la main d'œuvre jusque là non qualifiée. Par exemple dès 1975, deux nouvelles classifications ont été créées par la Société française de mécanique : OPF (Opérateur Professionnel de Fabrication) et APF (Agent Professionnel de Fabrication), chacune correspondant à un niveau précis de qualification.²⁰

2.2.2. Ambiguïté OQ/Technicien : niveau de diplôme et salaire

L'exemple qui suit est beaucoup plus anecdotique que le précédent, mais il montre bien comment un libellé d'emploi peut se former dans un « jeu » assez subtil et complexe.

Un jeune interrogé dans Génération 98 a livré un intitulé très précis « *technicien installation et maintenance systèmes de sécurité* ». On peut imaginer les variations possibles autour de ce libellé. Dans un questionnaire auto-administré du recensement, l'enquêté pourrait très bien simplifier sa réponse et se déclarer « *technicien* » ou « *technicien sécurité* ». En mode téléphonique assisté par ordinateur, certains enquêteurs pourraient juger cet intitulé trop long car sa transcription littérale casse le rythme de l'entretien. L'intitulé est alors réduit et peut devenir aussi bien « *technicien installation sécurité* » que « *technicien maintenance sécurité* ».

Chacune de ces variantes aboutit à un code profession différent. « *Technicien* » sera classé dans la catégorie des techniciens (CS 47). « *Technicien sécurité* » nous renseigne à la fois sur la classification professionnelle et le secteur d'activité ; mais il peut y avoir hésitation entre deux codes car on ne sait pas s'il s'agit d'un technicien d'études ou de maintenance²¹. L'appellation « *Techniciens installation sécurité* » est ambiguë ; elle doit être rapprochée de l'intitulé standard « installateur en alarme », assimilé aux « électriciens qualifiés du bâtiment » (6311) ; on la classe dans la catégorie des ouvriers qualifiés (CS 63) et non pas dans celle des techniciens (CS 47). De son côté, le « *Technicien maintenance sécurité* » sera classé aisément avec « les techniciens de maintenance, dépannage, en électricité, électronique, automatisme » (4717).

Paradoxalement, l'intitulé de référence (le plus précis) est difficile à coder car notre technicien sécurité a indiqué une double fonction. Doit-on le coder comme un « technicien installation sécurité » (6311) ou comme un « technicien maintenance sécurité » (4717) ?

On peut certes reprocher à l'enquêté de ne pas avoir indiqué sa fonction principale. Mais en prolongeant l'entretien, on s'aperçoit que l'ambiguïté de sa déclaration provient d'une politique de gestion de l'entreprise conduisant à une désindexation salariale. En effet, les jeunes techniciens de cette entreprise sont bien polyvalents. D'ailleurs, le jeune se déclare « technicien » dans la classification socioprofessionnelle, manifestant ainsi un statut acquis avec son diplôme (BTS en électronique). Mais la feuille de paye vient rappeler que dans cette entreprise les jeunes techniciens ont un salaire d'installateurs en alarme c'est-à-dire d'ouvriers qualifiés.

Cet exemple montre que les difficultés et aléas de classement ne renvoient pas seulement à des « ratés » dans une chaîne de traitement ou à une inadéquation entre d'un côté, la nomenclature dans laquelle on entend classer les emplois et de l'autre côté, des « catégories spontanées ». Ils renvoient

¹⁹ On se référera à l'ouvrage de Stéphane Beaud et Michel Pialoux : « Retour sur la condition ouvrière - Enquête aux usines Sochaux-Montbéliard », Paris, Fayard 1999.

²⁰ D'après une note de travail écrite par Laurence Coutrot le 27 juin 2002 : « Le cas de la Française de Mécanique ».

²¹ Les deux codes possibles sont :

4713 – Techniciens d'études, essais, contrôles en électricité, électronique ou bien

4717 – Techniciens de maintenance, dépannage, en électricité, électronique, automatisme.

aussi à une négociation constante des statuts et des rémunérations. Les organisations du travail, en évoluant, forgent de nouveaux repères. Ainsi la polyvalence ou certaines fonctions (comme la maintenance, la fonction marketing) vont inévitablement bousculer les clivages et on peut penser que les jeunes « expérimentent » plus que les autres ces changements.

C'est avec cette nouvelle vision des choses que nous allons donc analyser comment les discordances (et le « flou statistique » qui en résulte) se répartissent sur le champ des positions sociales et des emplois.

2.3. La robustesse des catégories sociales à l'épreuve des incertitudes de classement

Tout d'abord, les aléas de classement des emplois entre les grandes catégories socioprofessionnelles reflètent-ils une moindre structuration de la société en « classes sociales » ?

Comme le pointe Laurence Coutrot (Sociétés Contemporaines 2002), la pertinence des groupes sociaux est sérieusement remise en cause. Dans les entreprises, certaines conventions collectives seraient en train de bousculer la hiérarchie des classes²². La catégorie ouvrière est supposée être la plus affectée. D'après certains sociologues, les jeunes se définiraient de moins en moins comme ouvriers (Olivier Schwartz).

Mais loin d'apporter de l'eau au moulin de cette remise en cause, l'analyse des discordances de classement²³ montre que les deux grands groupes sociaux situés aux extrémités de la hiérarchie sociale résistent bien ou en tout cas qu'ils restent bien « discernables ».

2.3.1. Des murs solides mais des cloisons amovibles pour la catégorie ouvrière

Ainsi, 93 % des emplois classés dans le groupe des « ouvriers » par SICORE se trouvent rangés dans le même groupe social par les codeuses. Dans les trois quarts des cas, l'individu s'est d'ailleurs lui-même positionné comme « ouvrier » (en réponse à la question sur la position professionnelle).

En revanche, la répartition de ces emplois entre les six catégories ouvrières de la nomenclature socioprofessionnelle en 42 postes est bien plus « aléatoire ». A ce niveau de détail, le taux de divergences entre les deux codifications atteint 20 %. Après expertise, 10 % des emplois restent d'ailleurs impossibles à classer de manière certaine à partir des informations disponibles (voir tableau).

Ce résultat n'est pas nouveau. Il rejoint ce qu'a relevé Alain Chenu sur la base d'une confrontation entre le recensement et l'enquête emploi de 1982 : il y a, pour le groupe des ouvriers, une certaine « sûreté de classement » en grande masse mais une difficulté à ranger leurs emplois dans les « divisions internes à cette grande masse ».

2.3.2. La catégorie des emplois de cadre bien délimitée

Le résultat concernant les jeunes cadres (et professions intellectuelles supérieures) est moins conforme aux analyses de même type précédemment menées sur l'enquête emploi ou le Recensement de Population. En effet, les contours de cette catégorie apparaissent ici assez robustes et en tout cas

²² Par exemple, la nouvelle convention collective de l'assurance prévoit ainsi une grille fonctionnelle unique pour les cadres et non les cadres.

²³ Tout comme l'analyse de Laurence Coutrot.

nettement moins sensibles à la méthode de codification de l'information recueillie, que les catégories « employés » et « professions intermédiaires ». Ce résultat peut être lié à la spécificité des emplois de cadre occupés par de jeunes débutants, plus diplômés que les autres cadres et plus concentrés sur les sphères techniques du secteur privé. Il est également possible que dans l'intitulé de leur profession, ils se réfèrent davantage à leur diplôme (« ingénieur », par exemple) plutôt qu'à leur position dans l'organisation du travail (cf. étude en 2002 de H. Eckert sur les jeunes ingénieurs).

Il est important de noter ici que de tous les groupes sociaux, c'est pour les emplois de « cadre » que concordent le plus souvent la « position professionnelle » déclarée et la catégorie sociale codifiée.

2.3.3. Un flou important pour les jeunes en « minorité sexuelle » dans une catégorie professionnelle

La prise en compte du genre permet d'affiner le repérage des espaces les plus rebelles à la codification (graphique 3 et tableau 5). Globalement, les emplois occupés par des jeunes femmes semblent aussi difficiles à classer en catégories socioprofessionnelles que ceux de leurs homologues masculins. Mais derrière cette apparente égalité, se cachent des écarts de sens opposés au sein du groupe des employés (où les divergences sont plus fréquentes pour les hommes) et au sein du groupe ouvrier (où les divergences sont plus fréquentes pour les femmes). Alors que les règles de codification ignorent le sexe des individus, tout se passe comme si l'affectation d'un individu sur une catégorie où le sexe opposé domine, jetait le trouble. Cela se confirme au niveau de la catégorie socioprofessionnelle (42) puisque du côté des catégories très masculines comme celles des ouvriers qualifiés (de type industriel ou artisanal) ou des « policiers et militaires », les divergences sur les emplois occupés par des femmes sont nettement plus fréquentes que sur ceux des hommes. Du côté de certaines catégories très féminisées, au contraire, le taux de divergences sur les emplois occupés par des hommes dépasse celui des femmes et 61 % des hommes classés en « employés administratifs » se retrouvent classés ailleurs par l'équipe de codeurs.

On aurait pu penser que ce résultat était lié à l'attitude des codeuses qui « naturellement » (et contrairement à un automate « sans état d'âme ») auraient tendance à remettre en cause tout résultat « non conforme » (jugé douteux) et donc à réviser tout ce qui relève de l'« exception statistique ». Une telle pratique conduirait à biaiser les résultats en réduisant le taux de féminité au sein des catégories les plus masculines et en augmentant ce taux au sein des catégories les plus féminines. Or la comparaison de ces taux d'un codage à l'autre ne fait pas apparaître un tel biais (tableau 6). Le taux de féminité dans presque toutes les catégories féminines d'employés est même un peu moins élevé dans le résultat du codage manuel que dans celui de SICORE. Cela nous renvoie donc à d'autres explications possibles. On peut penser que lorsqu'un individu occupe un emploi appartenant à une catégorie de sexe opposé, il ne va pas sur les postes les plus typiques de cette catégorie et que les libellés correspondant entrent dans un champ mal balisé par la nomenclature.

Graphique 3 - Ecart entre les taux de divergence masculins et féminins sur les catégories socio-professionnelles (42)

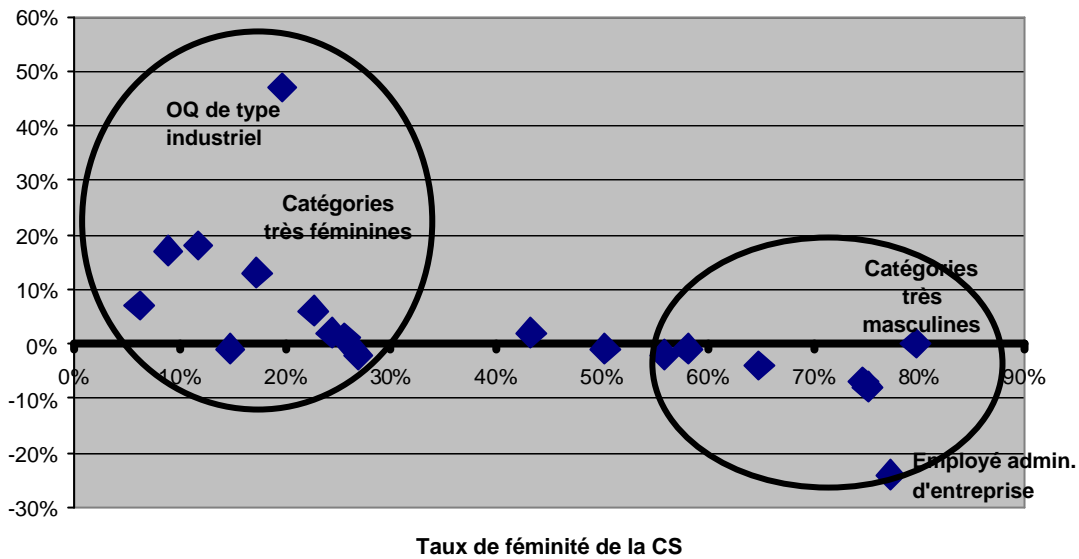


Tableau 5 : Fréquence des divergences PCS par sexe et groupe social

	Groupe social	Effectif	Taux de divergence du code profession		
			GS	CS	PCS
Ensemble	1-CADRE PROF.SUP	1672	6%	15%	33%
	2-PROF.INTERMEDIAIRE	2457	10%	21%	39%
	3-EMPLOYE	2273	11%	18%	31%
	4-OUVRIER	3449	7%	20%	33%
	5-AUTRES	149	33%	47%	65%
	Ensemble	10000	9%	19%	35%
Hommes	1-CADRE PROF.SUP	920	6%	14%	36%
	2-PROF.INTERMEDIAIRE	1153	12%	22%	45%
	3-EMPLOYE	602	17%	24%	37%
	4-OUVRIER	2885	5%	18%	31%
	5-AUTRES	117	29%	46%	63%
	Ensemble	5677	9%	20%	36%
Femmes	1-CADRE PROF.SUP	752	8%	16%	30%
	2-PROF.INTERMEDIAIRE	1304	9%	20%	34%
	3-EMPLOYE	1671	9%	16%	29%
	4-OUVRIER	564	13%	26%	40%
	5-AUTRES	32	46%	49%	71%
	Ensemble	4323	10%	19%	32%

Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

Tableau 6 : Taux de féminité des groupes sociaux suivant le codage PCS

GROUPE SOCIAL	Part des femmes		
	Codage SICORE	Codage Manuel	Ecart
1-CADRE PROF.SUP	43,5%	43,7%	0,2%
2-PROF.INTERMEDIAIRE	55,2%	56,5%	1,4%
3-EMPLOYE	74,4%	73,6%	-0,8%
4-OUVRIER	18,9%	17,9%	-1,0%
5-AUTRES	22,1%	18,7%	-3,5%
Ensemble	46,9%	46,9%	0,0%

Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

2.4. La robustesse des catégories sociales à l'épreuve des incertitudes de classement

Si l'examen des taux de divergence par grands groupes sociaux fait apparaître deux grandes « catégories » dont les délimitations semblent plus « claires » que les autres, il ne révèle pas de larges champs où le flou sur le code profession (fin) est beaucoup plus fréquent qu'ailleurs. C'est pourquoi nous allons maintenant prendre en compte des données concernant l'« environnement » des emplois occupés.

Une analyse des discordances dans le codage par secteur d'activité, par type d'employeur et de contrat de travail permet de repérer des « espaces économiques » nettement plus « rebelles » à la codification PCS que d'autres.

Les principaux résultats concernant les disparités sectorielles sont surprenants (graphique 4 et tableau 7).

Graphique 4 - Fréquence des divergences par secteur d'activité

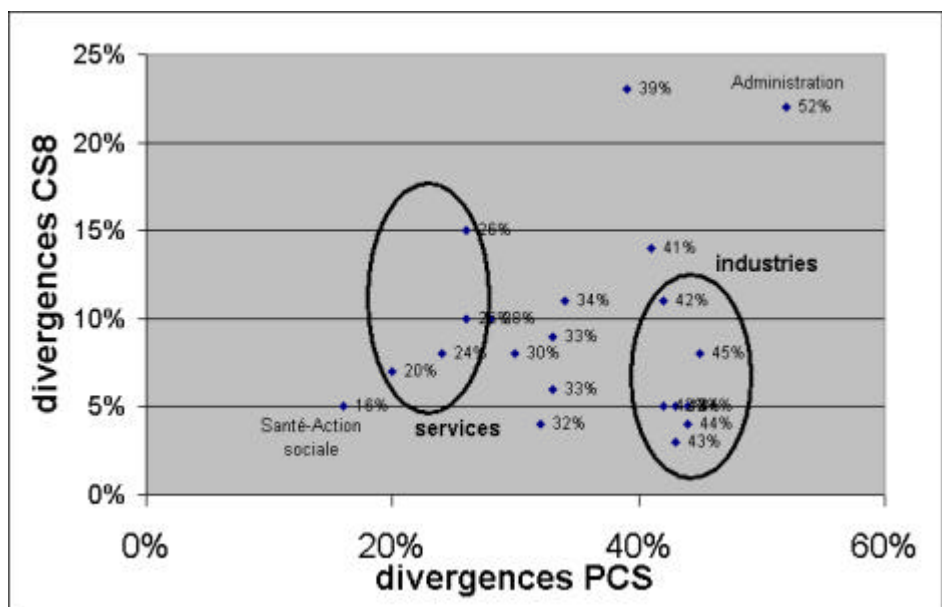


Tableau 7 : Fréquence des divergences PCS par secteur d'activité

Secteurs d'activité	Taux de divergence au niveau du code profession :	Effectif

	GS	CS	PCS	
Santé Action sociale	16%	11%	5%	915
Hôtels et restaurants	24%	20%	11%	558
Agriculture, sylviculture, pêche	26%	16%	10%	261
Services opérationnels	27%	20%	8%	183
Autres services aux particuliers	27%	22%	10%	865
Commerce et réparation automobile	29%	21%	9%	194
Commerce de détail	30%	25%	17%	224
Industries agricoles et alimentaires	32%	17%	10%	929
Education	32%	22%	6%	417
Construction	34%	22%	6%	680
Transports	38%	24%	20%	257
Activités financières et immobilières	38%	22%	11%	197
Autres Services aux entreprises	41%	25%	12%	841
Industries des biens de consommation	42%	17%	7%	318
Industries des biens intermédiaires	43%	10%	4%	340
Industrie automobile	43%	17%	7%	683
Energie	46%	18%	9%	396
Industries des biens d'équipements	49%	13%	4%	226
Commerce de gros, intermédiaires	51%	27%	8%	79
Administration	54%	32%	23%	506
Ensemble	35%	19%	9%	9069

Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

2.4.1. Incertitudes de classement fortes dans le secteur public

Malgré les grilles de classification statutaire du public, le secteur de l'administration cumule les discordances de codification au niveau fin comme au niveau des grands groupes sociaux. Aussi bien du côté de l'Etat que des Collectivités Territoriales, plus de la moitié des emplois ne trouvent pas le même classement en PCS et un quart des emplois n'aboutissent pas dans le même groupe social.

2.4.2. Un univers industriel mal balisé au-delà des grandes catégories ?

Autre surprise dans le palmarès : les secteurs industriels (y compris les plus traditionnels et ceux où la grande entreprise domine) apparaissent, eux aussi, parmi les activités économiques où la codification fine est la plus instable. Ainsi, dans les industries de biens d'équipement, biens de consommation, biens intermédiaires, dans l'automobile et l'énergie, 40 à 50 % des emplois se voient attribuer une PCS différente suivant le mode de codification. Certes, tous ces secteurs restent encore relativement structurés par les conventions collectives qui les couvrent. De ce fait, les taux de divergence à des niveaux plus regroupés de la nomenclature restent assez faibles (sauf pour l'énergie). Les affectations entre techniciens, ouvriers qualifiés et non qualifiés sont ici peu tributaires de la méthode. Mais tout se passe comme si un découpage plus fin des emplois relevait d'arbitrages subtils basés sur des informations annexes à défaut d'intitulés d'activité normalisés.

Paradoxalement, les secteurs des services (hôtellerie-restauration, services opérationnels, commerce et réparation de l'automobile, notamment mais aussi des secteurs modernes comme le conseil-assistance informatique) semblent plutôt bien « quadrillés » par la nomenclature PCS. Evidemment, le secteur de la santé et de l'action sociale, qui abrite des professions dont l'exercice est réglementé, est celui qui pose le moins de problèmes de codifications...

Face à un ensemble de résultats assez inattendus, il convient de rappeler que l'étude ne porte pas sur une photographie générale de l'ensemble de la population active mais sur un flux d'emplois correspondants aux expériences professionnelles de jeunes débutant dans la vie active. Dès lors, dans quelle mesure la spécificité de cette population intervient-elle ?

Par exemple, le statut précaire, avec lequel les jeunes sont souvent embauchés, contribue-t-il à maintenir une faible « affiliation professionnelle » et à se situer de façon floue dans l'organisation où ils travaillent ?

2.4.3. Dans le privé, la précarité d'emploi n'engendre pas une précarité du classement

Là encore, les premiers résultats viennent partiellement contredire l'intuition (tableau 8). Dans le secteur privé, toutes choses égales par ailleurs, les emplois en CDD ne font pas l'objet de taux de divergence plus élevés que les emplois occupés en CDI.

2.4.4. De nombreuses positions hors cadre, à l'origine du flou dans le public

En réalité, c'est surtout dans le secteur public que le statut joue un rôle déterminant sur la clarté dans le positionnement et la codification de la catégorie professionnelle. Les jeunes employés comme vacataires, sous CDD ou avec un contrat aidé sont nombreux. Ces contrats introduisent des positions « hors norme » qui induisent un aléa de codification important. Ainsi, les « emplois jeunes » ou les Contrats Emplois Solidarité font apparaître des intitulés d'emploi cumulant allègrement les néologismes²⁴... Du côté de l'Etat, le taux de divergence atteint 56 % pour les non-fonctionnaires contre seulement 13 % pour les fonctionnaires. Du côté des collectivités territoriales et hôpitaux, le contraste est un peu moins fort.

L'analyse par nature d'employeur et contrat vient donc ici éclairer une partie des disparités sectorielles. Le résultat d'ensemble sur le secteur administration s'explique en réalité par l'importance des recrutements de jeunes n'ayant pas le statut de fonctionnaire. Il est sans aucun doute spécifique à des jeunes entrant dans la vie active puisque le taux de contractuels et stagiaires chez les employeurs publics est faible au-delà de 30 ans.

Tableau 8 : Effet du type de contrat de travail sur la fréquence des divergences PCS

Secteur	Type de contrat	Taux de divergence au niveau du code profession			Effectif
		GS	CS	PCS	
<i>Privé</i>	CDI	8%	17%	36%	2553
	CDD	8%	16%	29%	2418
	interim	7%	17%	35%	1303
	contrats d'alternance	8%	16%	32%	302
	autres contrats aidés	7%	24%	37%	126
	Ensemble	8%	17%	33%	6707
<i>CT, Hopitaux</i>	fonctionnaire	2%	5%	12%	56
	CDI	6%	7%	15%	65
	CDD	13%	18%	26%	431

²⁴ Ces nouvelles appellations du type « ambassadeur du tri » ne sont pas toujours reconnues par SICORE. Lorsque c'est le cas, il s'agit fréquemment d'une confusion en écho avec un libellé existant dans les bases pré-existantes et correspondant à un autre métier.

	emploi jeune	27%	37%	61%	40
	CES et autres C.A.	8%	15%	47%	23
	Ensemble	12%	18%	28%	651
<i>Etat</i>	fonctionnaire	9%	9%	12%	437
	CDI	14%	30%	34%	72
	CDD	19%	33%	47%	318
	emploi jeune	4%	78%	87%	89
	CES et autres C.A.	32%	53%	53%	19
	Ensemble	13%	31%	39%	962

Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

2.5. Professionnalisation de la formation initiale et sûreté de classement professionnel

On peut penser que l'existence de diplômes professionnels dont la spécialité est définie de façon plus ou moins précise, contribue à structurer et à baliser le champ des emplois, ou tout du moins peut jouer un rôle dans une certaine standardisation des appellations d'activités. On peut supposer que pour son titulaire, le diplôme vient renforcer l'identité professionnelle.

Certes comme nous l'avons vu précédemment, cette structuration de l'espace professionnel par des formations finalisées se vérifie bien dans le cas des formations aux métiers de la santé et de l'action sociale. A l'opposé, les discordances entre les deux codages culminent chez les jeunes sortant des études avec un DEUG en poche : un quart de leurs emplois se retrouvent même classés dans un groupe social différent.

Mais en dehors de ces deux extrêmes, avoir suivi une formation professionnelle ne semble pas accroître la sûreté de classement de son emploi dans la nomenclature de référence (tableau 9). Le taux de divergence sur le code profession est élevé pour les sortants de BTS ou DUT ainsi que pour les diplômés des écoles d'ingénieur ou de commerce alors qu'il est plus faible pour les diplômés de second cycle universitaire. De leur côté, les emplois occupés par les plus bas niveaux de formation ne semblent pas plus difficiles à positionner que les autres.

Ainsi, une forte professionnalisation de sa formation initiale ne semble pas conduire vers des énoncés d'emplois plus faciles à classer avec certitude dans la grille PCS. Mais on revient là sur une des questions de départ : l'emploi correspond-il à la formation ?

Tableau 9 : Fréquence des divergences par niveau de formation

Niveau et spécialité de formation	Taux de divergence au niveau du code profession :		
	GS	CS	PCS
01 non qualifiés	8%	20%	34%
02 cap ou bep non diplômés, 2de, 1re	7%	18%	33%
03 cap ou bep tertiaire	8%	16%	27%
04 cap ou bep industriel	6%	18%	29%
05 bac non diplômés	6%	17%	29%
06 bac tertiaire	9%	15%	24%
07 bac industriel	6%	22%	43%
08 bac+1, bac+2 non diplômés	12%	25%	39%
09 de la santé et du social	1%	2%	7%

10 DEUG, DEUST...	21%	34%	50%
11 BTS, DUT tertiaire	16%	25%	45%
12 BTS, DUT industriel	10%	21%	48%
13 2d cycle LSH, gestion	12%	20%	32%
14 2d cycle maths, sc, tech	12%	21%	41%
15 3eme cycle LSH, gestion	12%	21%	37%
16 écoles commerce	3%	20%	50%
17 3eme cycle maths, sc, tech	4%	14%	34%
18 écoles ingénieur	5%	10%	46%

Source : CEREQ - Enquête Génération 1998 - Interrogation 2001

3. Conclusion

Cette investigation méthodologique avait une double ambition :

- évaluer jusqu'où le mode de codification de la profession pouvait faire varier les conclusions d'une analyse statistique du lien formation-emploi basée sur l'exploitation d'une enquête sur les cheminements en début de vie active telle que Génération 98,
- mieux comprendre l'origine des fréquentes divergences entre les résultats d'une procédure manuelle et ceux d'une procédure automatisée dans la codification en PCS des emplois.

Sur le premier point, malgré l'ampleur des divergences (35 % de discordances sur le code PCS fin), nous avons montré que les indicateurs usuellement utilisés dans l'analyse du lien formation-emploi, sont peu sensibles au mode de codification dès lors qu'ils portent sur une population assez large. Néanmoins, sur certains champs, les fréquentes discordances méritent que les libellés en clair soient ré-examinés. D'autant plus qu'elles peuvent permettre de mieux comprendre ce qui se joue autour des énoncés d'emploi : introduction de la compétence, affirmation de nouvelles fonctions dans le poste...

Sur le deuxième point, il apparaît que seule la moitié des divergences a été classée par un « expert PCS » comme une erreur manifeste de l'homme ou de la machine (au regard du manuel officiel de codification) et que dans bien des cas, c'est la qualité du matériau recueilli lors de l'interrogation qui semble insuffisante ou inadaptée. Certaines dispositions techniques pourraient limiter encore la fréquence de libellés de profession trop imprécis. Mais plus fondamentalement, l'obtention de déclarations permettant une codification très fine de la profession dans la nomenclature en 455 postes ne relève-t-elle pas de l'acharnement thérapeutique ?

En effet, ce n'est pas tant le classement des jeunes entre les grands groupes sociaux qui pose problème. La codification à ce niveau semble assez stable, même si un certain flou apparaît entre professions intermédiaires et employés. En réalité, c'est surtout dans la codification fine que les procédés divergent.

Le flou de codification, observé à ce niveau, varie en fonction de l'environnement de l'emploi (secteur d'activité, type de contrat, etc.) mais rarement dans le sens attendu. Ainsi, dans les univers les plus structurés par les conventions collectives, la grande industrie ou l'administration, les taux de divergence sont assez élevés. Dans le secteur privé, le flou de codification n'est pas plus important pour les emplois en CDD que pour les emplois occupés en CDI. Le lien entre précarité de l'emploi et flou dans le positionnement social ne semble donc pas évident.

Au final, l'analyse de la répartition du flou de codification dans l'espace socio-économique met certes en lumière quelques micro-secteurs très concernés (comme les emplois précaires dans la fonction publique), mais le taux de divergence entre les deux modes de codifications reste élevé sur un champ très large. Il s'agit donc d'un phénomène assez général²⁵ et on peut se demander au final si la codification d'une PCS à quatre chiffres dans une enquête de ce type constitue une ambition raisonnable. La refonte récente de la nomenclature a incontestablement permis de prendre en compte quelques grandes mutations dans les organisations du travail et de solutionner plusieurs cas posant de sérieux problèmes. Des améliorations techniques sont possibles, allant de plus en plus vers une attribution de la PCS par l'enquêteur en cours d'entretien. Mais il faut maintenir une mesure de l'aléa de codage et si celui-ci reste important, ne faudra-t-il pas en tirer les conséquences et aller vers une nomenclature un peu plus regroupée ?

Bibliographie

- [1] Beaud S. et Pialoux M., « Retour sur la condition ouvrière - Enquête aux usines Sochaux-Montbéliard », Paris, Fayard, 1999.
- [2] Cart B. et Toutain M.H, « Quand les jeunes entrent dans l'emploi, Chapitre V – La maintenance industrielle : une ouverture pour les jeunes ? », *La dispute*, pp 115-126, juin 2002
- [3] Chenu. A., « La descriptibilité statistique des professions », *Sociétés Contemporaines*, n° 26, pp109-137, avril 1997
- [4] Chenu. A. et Guglielmetti F., « Coder la profession : nouvelles procédures, nouveaux enjeux », Paris, INSEE, VIIèmes Journées de Méthodologie Statistique, décembre 2000
- [5] Coutrot L., « Les catégories socio-professionnelles, Changement des conditions, permanence des positions ? », *Sociétés Contemporaines*, n° 45-46, pp1-23, 2002
- [6] Desrosières A., Thévenot L., « Les catégories socioprofessionnelles », Paris, *La découverte*, 1988
- [7] Eckert H., « Quand les jeunes entrent dans l'emploi, Chapitre III – Le titre ou la position ? De jeunes ingénieurs énoncent leur emploi », *La dispute*, pp 75-91, juin 2002
- [8] Eckert H. et Maillard D., « Quand les jeunes entrent dans l'emploi, Chapitre I – Commerciaux et vendeurs : des intitulés d'emploi aux groupes professionnels », *La dispute*, pp 43-57, juin 2002
- [9] Guglielmetti F., « L'appréhension de la profession et de la catégorie sociale dans une enquête par dépôt-retrait », Paris, INSEE, note 77/F420, juin 2000

²⁵ Encore faut-il préciser que le « flou de codification » analysé ici, était uniquement limité à une variabilité dans le mode de traitement de l'information recueillie. Une analyse sur les discordances qui résulteraient d'une double déclaration du même indivi du à deux moments séparés pourrait utilement prolonger ce travail.

[10] Kramarz F., « Déclarer sa profession », *Revue française de sociologie*, n° XXXII-1, pp 3-27, janvier-mars 1991

[11] Schwartz O., « Le monde privé des ouvriers, hommes et femmes du Nord », Paris, *PUF*, 1990